Electronic instrumentation

VISIT…

LANZAROTE
Caliente.COM

Electronic instrumentation

P.P.L. Regtien

VSSD

© **VSSD**

Second edition 2005

Published by:
VSSD
Leeghwaterstraat 42, 2628 CA Delft, The Netherlands
tel. +31 15 27 82124, telefax +31 15 27 87585, e-mail: hlf@vssd.nl
internet: http://www.vssd.nl/hlf
URL about this book: **http://www.vssd.nl/hlf/e008.htm**
A collection of digital pictures and/or an electronic version can be made available for lecturers who adopt this book. Please send a request by e-mail to hlf@vssd.nl

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher.

ISBN 90-71301-43-5
NUR 959
Keywords: electronic instrumentation

Preface

Electronic systems have made deep inroads into every aspect of daily life. One need only look around homes, offices and industrial plants to see that they feature almost everywhere. Indeed, it is practically impossible to name any appliances, tools or instruments that do not contain electronic components. In order to compete with rival companies or just remain a step ahead of them, the designers of technical systems and innovative products must be fully aware of both the assets and the limitations of electronic components and systems. Users of electronic systems also need to have a basic knowledge of electronic principles. In order to fully exploit an instrument's potential, to be aware of its limitations, to correctly interpret the measurement results and to be able to arrive at well-balanced decisions relating to the purchasing, repairing, expansion or replacement of electronic equipment, all users of such systems also need to have a basic knowledge of electronic principles.

This book offers such basic knowledge and provides guidance on how to obtain the relevant skills. The kinds of topics dealt with are operating principles, the performance of analog and digital components and circuits, and the precise characteristics of electronic measuring systems. Throughout the book, every endeavor is made to impart a critical attitude to the way in which such instruments should be implemented.

The book is based on various series of courses on electronics and electronic instrumentation that were given by the author during the many years when he lectured at Delft University of Technology in the Netherlands. The courses were designed for students from various departments such as: Mechanical Engineering, Aeronautical Engineering and Mining Engineering. When numbers of non-Dutch-speaking Master of Science students started to rise it became necessary to publish an English version of the book.

The particular way in which the book has been organized makes it suitable for a much wider readership. To meet the demands of divergent groups it has been structured in a modular fashion. Each chapter discusses just one particular topic and is divided into two parts: the first part provides the basic principles while more specific information is given in the second part. Each chapter ends with a summary and several exercises. Answers to all the exercises are given at the back of the book. This approach is conducive to self-study and to the composition of tailor-made course programs.

The required background knowledge is a basic grounding in mathematics and physics equivalent to any first-year academic level. No background knowledge of electronics is needed to understand the contents of the book. For further information on particular subjects the reader is referred to the many course books that exist on the subjects of electronics, measurement techniques and instrumentation.

I am indebted to all the people who contributed to the realization of this book. In particular I would like to thank Johan van Dijk who carefully refereed the original Dutch text. I am grateful also to Reinier Bosman for working out all the exercises, to G. van Berkel for creating the more than 600 illustrations, to Jacques Schievink for processing the original Dutch editions and this English version of the book and to Diane Butterman for reviewing the entire English text.

Paul Regtien
Hengelo, August 2004

Contents

PREFACE V

1 MEASUREMENT SYSTEMS
- 1.1 System functions 1
- 1.2 System specifications 6
- SUMMARY 12
- EXERCISES 13

2 SIGNALS
- 2.1 Periodic signals 14
  - 2.1.1 A classification of signals 14
  - 2.1.2 Signal values 15
  - 2.1.3 Signal spectra 17
- 2.2 Aperiodic signals 22
  - 2.2.1 Complex Fourier series 22
  - 2.2.2 The Fourier integral and the Fourier transform 24
  - 2.2.3 A description of sampled signals 26
  - 2.2.4 A description of stochastic signals 27
- SUMMARY 32
- EXERCISES 34

3 NETWORKS
- 3.1 Electric networks 37
- 3.2 Generalized network elements 42
- SUMMARY 46
- EXERCISES 47

4 MATHEMATICAL TOOLS
- 4.1 Complex variables 50

4.1.1 The properties of complex variables 50
4.1.2 The complex notation of signals and transfer functions 52
4.1.3 Impedances 52
4.2 Laplace variables 54
4.2.1 The Laplace transform 54
4.2.2 Solving differential equations with the Laplace transform 56
4.2.3 Transfer functions and impedances in the p-domain 57
4.2.4 The relation to the Fourier integral 59
SUMMARY 59
EXERCISES 60

5 MODELS

5.1 System models 63
5.1.1 Two-terminal networks 63
5.1.2 Two-port networks 64
5.1.3 Matching 68
5.1.4 Decibel notation 71
5.2 Signal models 72
5.2.1 Additive errors 72
5.2.2 Noise 75
SUMMARY 76
EXERCISES 78

6 FREQUENCY DIAGRAMS

6.1 Bode plots 80
6.1.1 First order systems 80
6.1.2 Higher order systems 82
6.2 Polar plots 85
6.2.1 First order functions 85
6.2.2 Higher order functions 87
SUMMARY 89
EXERCISES 91

7 PASSIVE ELECTRONIC COMPONENTS

7.1 Passive circuit components 93
7.1.1 Resistors 93
7.1.2 Capacitors 95
7.1.3 Inductors and transformers 97
7.2 Sensor components 100
7.2.1 Resistive sensors 100
7.2.2 Inductive sensors 104
7.2.3 Capacitive sensors 106
7.2.4 Thermoelectric sensors 107
7.2.5 Piezoelectric sensors 110

SUMMARY 112
EXERCISES 113

8 PASSIVE FILTERS
8.1 First and second order RC-filters 116
8.1.1 Low-pass first-order RC-filter 116
8.1.2 Highpass first-order RC-filter 119
8.1.3 Bandpass filters 122
8.1.4 Notch filters 123
8.2 Filters of higher order 123
8.2.1 Cascading first-order RC-filters 124
8.2.2 Approximations of the ideal characteristics 125
SUMMARY 127
EXERCISES 128

9 PN-DIODES
9.1 The properties of pn-diodes 130
9.1.1 The operation of pn-diodes 130
9.1.2 Photodiodes 134
9.1.3 Light-emitting diodes (LEDs) 136
9.2 Circuits with pn-diodes 137
9.2.1 Limiters 137
9.2.2 Peak detectors 139
9.2.3 Clamp circuits 140
9.2.4 DC voltages sources 143
SUMMARY 144
EXERCISES 145

10 BIPOLAR TRANSISTORS
10.1 The properties of bipolar transistors 148
10.1.1 Construction and characteristics 148
10.1.2 Signal amplification 151
10.2 Circuits with bipolar transistors 152
10.2.1 Voltage-to-current converter 152
10.2.2 The voltage amplifier stage with base-current bias 154
10.2.3 The voltage amplifier stage with a base-voltage bias 157
10.2.4 The emitter follower 160
10.2.5 The differential amplifier stage 162
SUMMARY 163
EXERCISES 164

11 FIELD-EFFECT TRANSISTORS
11.1 The properties of field-effect transistors 167
11.1.1 Junction field-effect transistors 167

11.1.2 MOS field-effect transistors 171
11.2 Circuits with field-effect transistors 174
11.2.1 Voltage-to-current converter 174
11.2.2 The voltage amplifier stage 174
11.2.3 The source follower 176
11.3 SUMMARY 177
EXERCISES 178

12 OPERATIONAL AMPLIFIERS

12.1 Amplifier circuits with ideal operational amplifiers 182
12.1.1 Current-to-voltage converters 184
12.1.2 Inverting voltage amplifiers 185
12.1.3 Non-inverting voltage amplifiers 186
12.1.4 Differential amplifiers 186
12.1.5 Instrumentation amplifiers 188
12.2 Non-ideal operational amplifiers 189
12.2.1 The specifications of operational amplifiers 189
12.2.2 Input offset voltage 190
12.2.3 Finite voltage gain 193
SUMMARY 195
EXERCISES 196

13 FREQUENCY SELECTIVE TRANSFER FUNCTIONS WITH OPERATIONAL AMPLIFIERS

13.1 Circuits for time domain operations 198
13.1.1 The integrator 199
13.1.2 Differentiator 202
13.1.3 Circuits with PD, PI and PID characteristics 203
13.2 Circuits with high frequency selectivity 205
13.2.1 Resonance filters 205
13.2.2 Active Butterworth filters 210
SUMMARY 211
EXERCISES 212

14 NONLINEAR SIGNAL PROCESSING WITH OPERATIONAL AMPLIFIERS

14.1 Nonlinear transfer functions 215
14.1.1 Voltage comparators 215
14.1.2 Schmitt-trigger 217
14.1.3 Voltage limiters 219
14.1.4 Rectifiers 221
14.2 Nonlinear arithmetic operations 222
14.2.1 Logarithmic converters 222
14.2.2 Exponential converters 225

14.2.3 Multipliers 225
14.2.4 Other arithmetic operations 227
14.2.5 A piecewise linear approximation of arbitrary transfer functions 229
SUMMARY 231
EXERCISES 232

15 ELECTRONIC SWITCHING CIRCUITS

15.1 Electronic switches 235
15.1.1 The properties of electronic switches 235
15.1.2 Components as electronic switches 239
15.2 Circuits with electronic switches 243
15.2.1 Time multiplexers 243
15.2.2 Sample-hold circuits 245
15.2.3 Transient errors 248
SUMMARY 251
EXERCISES 252

16 SIGNAL GENERATION

16.1 Sine wave oscillators 255
16.1.1 Harmonic oscillators 255
16.1.2 Harmonic oscillator circuits 258
16.2 Voltage generators 261
16.2.1 Triangle voltage generators 261
16.2.2 The ramp generator 263
16.2.3 Square wave and pulse generators 265
16.2.4 Voltage-controlled oscillators 266
SUMMARY 267
EXERCISES 268

17 MODULATION AND DEMODULATION

17.1 Amplitude modulation and demodulation 273
17.1.1 Theoretical background 273
17.1.2 Amplitude modulation methods 275
17.1.3 Demodulation methods 279
17.2 Systems based on synchronous detection 281
17.2.1 The phase-locked loop 282
17.2.2 Lock-in amplifiers 284
17.2.3 Chopper amplifiers 284
SUMMARY 286
EXERCISES 287

18 DIGITAL-TO-ANALOGUE AND ANALOGUE-TO-DIGITAL CONVERSION
18.1 Parallel converters 289
18.1.1 Binary signals and codes 289
18.1.2 Parallel DA-converters 292
18.1.3 Parallel AD-converters 296
18.2 Special converters 299
18.2.1 The serial DA-converter 300
18.2.2 The direct AD converter 301
18.2.3 Integrating AD-converters 302
SUMMARY 305
EXERCISES 306

19 DIGITAL ELECTRONICS
19.1 Digital components 308
19.1.1 Boolean algebra 308
19.1.2 Digital components for combinatory operations 313
19.1.3 Digital components for sequential operations 316
19.1.4 The SR flip-flop 316
19.1.5 JK flip-flops 318
19.2 Logic circuits 320
19.2.1 Digital multiplexer 320
19.2.2 The digital adder 321
19.2.3 Digital counters. 323
19.2.4 Shift registers 325
19.2.5 An application example 328
SUMMARY 333
EXERCISES 334

20 MEASUREMENT INSTRUMENTS
20.1 Stand-alone measurement instruments 336
20.1.1 Multimeters 337
20.1.2 Oscilloscopes 337
20.1.3 Signal generators 343
20.1.4 Counters, frequency meters and time meters 344
20.1.5 Spectrum analyzers 345
20.1.6 Network analyzers 345
20.1.7 Impedance analyzers 347
20.2 Computer-based measurement instruments 347
20.2.1 Bus structures 348
20.2.2 An example of a computer-based measurement system 351
20.2.3 Virtual instruments 354
SUMMARY 355
EXERCISES 356

21 MEASUREMENT UNCERTAINTY
21.1 Measurement uncertainty described 358
21.1.1 Types of uncertainty 358
21.1.2 Error propagation 361
21.2 Measurement interference 362
21.2.1 Causes of interference 363
21.2.2 Remedies 365
SUMMARY 369
EXERCISES 370

APPENDIX 373

ANSWERS TO EXERCISES download from http://www.vssd.nl/hlf/e008.htm

INDEX 391

1 Measurement systems

The aim of any *measuring system* is to obtain information about a physical process and to find appropriate ways of presenting that information to an observer or to other technical systems. With electronic measuring systems the various instrument functions are realized by means of electronic components.

Various basic system functions will be introduced in the first part of this chapter. The extent to which an instrument meets the specified requirements is indicated by the system specifications, all of which will be discussed in the second part of the chapter.

1.1 System functions

A measuring system may be viewed as a transport channel for the exchanging of information between measurement objects and target objects. Three main functions may be distinguished: data acquisition, data processing and data distribution (Figure 1.1).

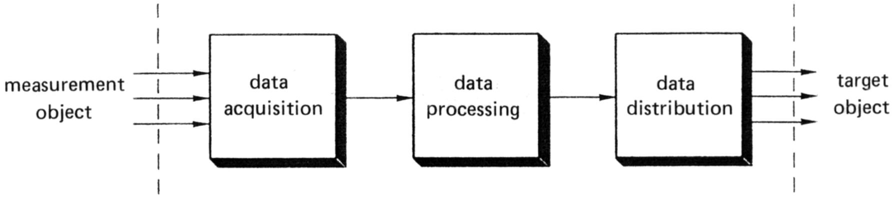


Figure 1.1. The three main functions of any measuring system.

- *Data acquisition*: this involves acquiring information about the measurement object and converting it into electrical measurement data. What multiple input, as illustrated in Figure 1.1, indicates is that invariably more than one phenomenon may be measured or that different measurements may be made, at different points, simultaneously. Where there are single data outputs this means that all data is transferred to the next block through a single connection.
- *Data processing*: this involves the processing, selecting or manipulating – in some other way – of measurement data according to a prescribed program. Often a processor or a computer is used to perform this function.

- *Data distribution*: the supplying of measurement data to the target object. If there is multiple output then several target instruments may possibly be present, such as a series of control valves in a process installation.

It should be pointed out that the above subdivision cannot always be made; part of the system may sometimes be classified as both data acquisition and data processing. Some authors call the entire system shown in Figure 1.1 a data acquisition system, claiming that the data is not obtained until the target object is reached.

In the next section the data acquisition and data distribution parts are subdivided into smaller functional units.

Since most physical measurement quantities are non-electric, they should first be converted into an electrical form in order to facilitate electronic processing. Such conversion is called transduction and it is effected by a transducer or sensor (Figure 1.2). In general, the transducer is kept separate from the main instrument and can be connected to it by means of a special cable.

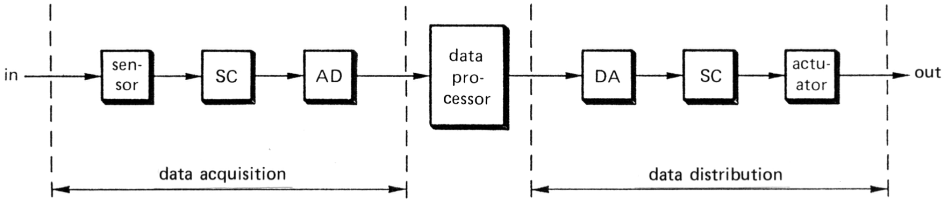


Figure 1.2. A single channel measuring system.

The sensor or input transducer connects the measuring system to the measurement object; it is the input port of the system through which the information enters the instrument.

Many sensors or transducers produce an analog signal; that is a signal whose value, at any given moment, is a measure of the quantity to be measured: the signal continuously follows the course of the input quantity. However, much of the processing equipment can only cope with digital signals, which are binary coded signals. A digital signal only contains a finite number of distinguishable codes, usually a power of 2 (for instance $2^{10} = 1024$).

The analog signal must be converted into a digital signal. This process is known as analog-to-digital conversion or, AD-conversion. Analog-to-digital conversion comprises three main processes, the first of which is sampling where, at discrete time intervals, samples are taken from the analog signal. Each sampled value is maintained for a certain time interval, during which the next processes can take place. The second step is quantization. This is the rounding off of the sampled value to the nearest of a limited number of digital values. Finally, the quantized value is converted into a binary code.

Both sampling and quantization may give rise to loss of information. Under certain conditions, though, such loss can be limited to an acceptable minimum.

The output signal generated by a transducer is seldom suitable for conversion into a digital signal, the converter input should first satisfy certain conditions. The signal processing required to fulfill such conditions is termed signal conditioning. The various processing steps required to achieve the proper signal conditions will be explained in different separate chapters. The main steps, however, will be briefly explained below.

- *Amplification*: in order to increase the signal's magnitude or its power content.
- *Filtering*: to remove non-relevant signal components.
- *Modulation*: modification of the signal shape in order to enable long-distance signal transport or to reduce the sensitivity to interference during transport.
- *Demodulation*: the reverse process operation to modulation.
- *Non-linear* and *arithmetical operations*: such as logarithmic conversion and the multiplication of two or more signals.

It goes without saying that none of the above operations should affect the information content of the signal.

After having been processed by the (digital) processor, the data are subjected to a reverse operation (Figure 1.2). The digital signal is converted into an analog signal by a digital-to-analog or DA converter. It is then supplied to an actuator (alternative names for this being: effector, excitator and output transducer), which transforms the electrical signal into the desired non-electric form. If the actuator cannot be connected directly to the DA converter, the signal will first be conditioned. This conditioning usually involves signal amplification.

The actuator or output transducer connects the measurement system to the target object, thus becoming the instrument’s output port through which the information leaves the system.

Depending on what is the goal of the measurement, the actuator will perform various functions such as, for instance: *indicating* by means of a digital display; *registering* (storing) with such things as a printer, a plotter or a magnetic disk; or *process controlling* with the aid of a valve, a heating element or an electric drive.

The diagram given in Figure 1.2 refers only to one input variable and one output variable. For the processing of more than one variable, one could take a set of single channel systems. Obviously this is neither efficient nor necessary. The processor shown in Figure 1.2, in particular, is able to handle a large number of signals, thanks to its high data processing speed. Figure 1.3 gives the layout of a multi-channel measuring system that is able to handle multiple inputs and outputs using only one (central) processor.

Central processing of the various digital signals can be effected by means of multiplexing. The digital multiplexer denoted in Figure 1.3 connects the output of each AD converter to the processor in an alternating fashion. The multiplexer may be viewed as an electronically controlled multi-stage switch, controlled by the processor. This type of multiplexing is called time multiplexing because the channels are scanned and their respective signals are successively transferred – in terms of time – to the processor. Another type of multiplexing, frequency multiplexing, will be discussed in a later section.

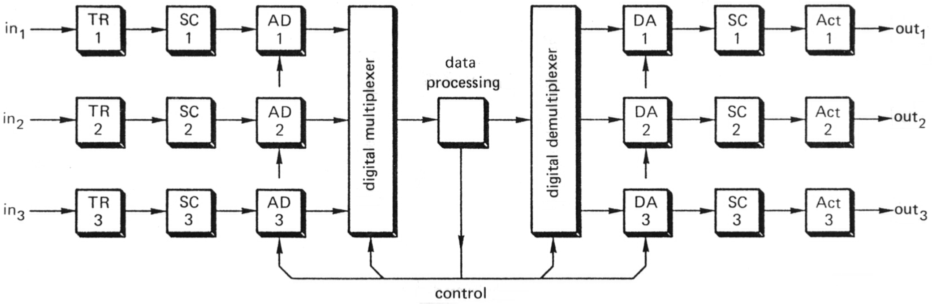


Figure 1.3. A three-channel measuring system with one central processor.
TR = transduction, SC = signal conditioning.

At first sight it would appear that the concept of time multiplexing has the disadvantage that only the data taken from the selected channel is processed while the information derived from the non-selected channels is blocked. It can be demonstrated that when the time between two successive selections for a particular channel is made sufficiently short the information loss will be negligible. An explanation of what precisely is meant by “sufficiently short” will be given in Section 2.2.

Figure 1.3 clearly shows that a system with many sensors or actuators will also contain large numbers of signal processing units, thus making it expensive. In such cases the principle of multiplexing can also be applied to the AD and DA converters. Figure 1.4 shows the layout of such a measurement system in which all the conditioned signals are supplied to an analog multiplexer. It is even possible to have a central signal conditioner placed behind the multiplexer so as to further reduce the number of system components. It is possible to extend the process of centralizing instrument functions to the data distribution part of the system. An analog multiplexer distributes the converted analog signals over the proper output channels. It is not common practice for output signal conditioners to be multiplexed because multiplexers are not usually designed to deal with large power signals.

Although the functions of analog and digital multiplexers are similar, their design is completely different. Digital multiplexers only deal with digital signals which have better noise and interference immunity than analog signals. Digital multiplexers are therefore far less critical (and less expensive) than analog multiplexers. The same goes for the AD converters. In Figure 1.3 it can be seen that each AD converter has a full multiplexer cycle period in which to perform a conversion. In the system shown in Figure 1.4, the conversion ought to be completed within the very short period of time when a channel is connected to the processor. This system configuration thus requires a high speed (and a higher priced) converter. The centralized system contains a reduced number of more expensive components. Whether one opts for a centralized or a distributed system will depend very much on the number of channels.

In certain situations the measurement signals and control signals have to be transported over long distances. This instrumentation aspect is known as telemetry. A telemetry channel consists of an electric conductor (for instance a telephone cable), an optical link (like a glass fiber cable) or a radio link (e.g. made via a communication satellite). To reduce the number of lines, which are invariably expensive, the concept of multiplexing is used (Figure 1.5). Instead of time multiplexing, telemetry systems use frequency multiplexing. Each measurement signal is converted to a frequency band assigned to that particular signal. If the bands do not overlap, the converted signals can be transported simultaneously over a single transmission line. When they arrive at the desired destination the signals are demultiplexed and distributed to the proper actuators. More details on this type of multiplexing will be given elsewhere in this book.

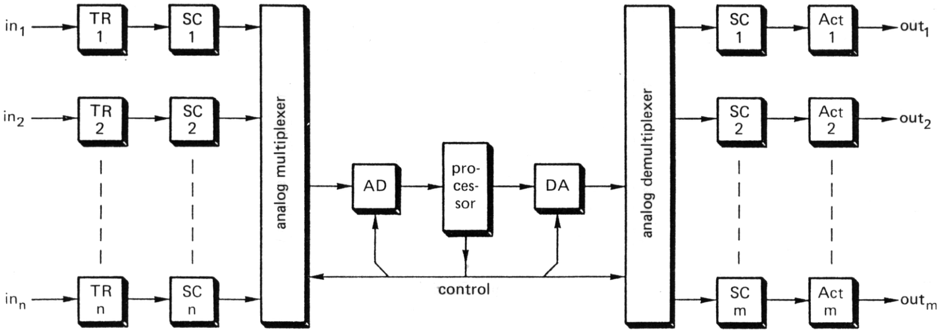

in1
TR 1
SC 1
in2
TR 2
SC 2
inn
TR n
SC n
analog multiplexer
AD
pro-ces-sor
DA
control
analog demultiplexer
SC 1
Act 1
out1
SC 2
Act 2
out2
SC m
Act m
outm

Figure 1.4. A multi-channel measuring system with a centralized processor and AD and DA-converters. For an explanation of the abbreviations see Figure 1.3.

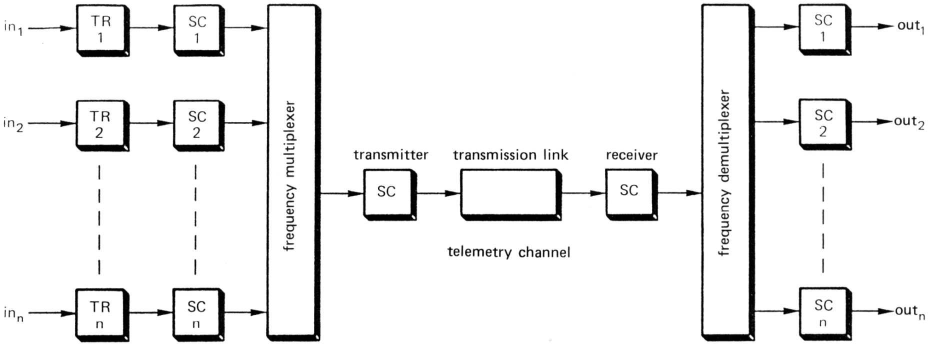

in1
TR 1
SC 1
in2
TR 2
SC 2
inn
TR n
SC n
frequency multiplexer
transmitter
transmission link
receiver
SC
SC
telemetry channel
frequency demultiplexer
SC 1
out1
SC 2
out2
SC n
outn

Figure 1.5. A multi-channel measuring system with frequency multiplexing.

Signals can be transmitted in analog or digital form. Digital transport is preferable if high noise immunity is required, for instance for very long transport channels or links that pass through a noisy environment.

1.2 System specifications

A measurement system is designed to perform measurements according to the relevant specifications. Such specifications convey to the user of the instrument to what degree the output corresponds with the input. The specifications reflect the quality of the system.

The system will function correctly if it meets the specifications given by the manufacturer. If that is not the case it will fail, even if the system is still functioning in the technical sense. Any measuring instrument and any subsystem accessible to the user has to be fully specified. Unfortunately, many specifications lack clarity and completeness.

The input signal of the single channel system given in Figure 1.6 is denoted as x and its output signal as y. The relationship between x and y is called the system transfer.

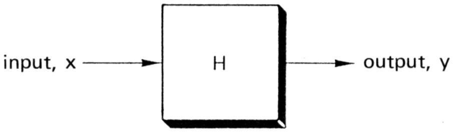


Figure 1.6. Characterization of a system with input *x*, output *y* and transfer *H*.

By observing the output, the user is able to draw conclusions about the input. The user therefore has to be completely familiar with the system's transfer. Deviations in the transfer may cause uncertainties about the input and so result in measurement errors. Such deviations are permitted, but only within certain limits which are the tolerances of the system. Those tolerances also constitute part of the specifications. In the following pages the main specifications of a measurement system will be discussed.

The user should first of all be familiar with the operating range of the system. The operating range includes the measurement range, the required supply voltage, the environmental conditions and possibly other parameters.

Example 1.1

A manufacturer of a digital temperature-measuring instrument gives the following description of the operating range:

* *measuring range: -50°C to 200°C;*
* *permitted operational temperature: –10°C to 40°C;*
* *storage temperature: -20°C to 85°C*
* *mains voltage: 220 V ±15%, 50...60 Hz; can be switched to 115 V, 127 V, 240 V ± 15%, 50...60 Hz;*
* *analog outputs: 0-10 V (load > 2kΩ) and 0-20 mA or 4-20 mA (load < 600Ω).*

All other specifications only apply under the condition that the system has never before been taken beyond its permitted operating range.

The resolution indicates the smallest detectable change in input quantity. Many system parts show limited resolution. A few examples of this are these: a wire-wound potentiometer for the measurement of angles has a resolution set by the windings of the helix – the resistance between the slider and the helix changes leap-wise as it rotates; a display presenting a measurement value in numerals has a resolution equal to the least significant digit.

The resolution is expressed as the smallest detectable change in the input variable: Δx_{min}. Sometimes this parameter is related to the maximum value x_{max} that can be processed, the so-called full-scale value or FS of the instrument, resulting in the resolution expressed as $\Delta x_{\mathrm{min}}/x_{\mathrm{max}}$ or $x_{\mathrm{max}}/\Delta x_{\mathrm{min}}$. This mixed use of definitions seems very confusing. However, it is easy to see from the units or the value itself which definition is used.

Example 1.2

The resolution of a four-digit decimal display with a fixed decimal point in the third position from the left is 0.1 units. The maximum indication apparently equals 999.9 units, which is about 1000. The resolution of this display is therefore 0.1 units or 10^{-4} *or* 10^{4}.

The inaccuracy is a measure of the total uncertainty of the measurement result that may be caused by all kinds of system errors. It comprises calibration errors, long and short-term instability, component tolerances and other uncertainties that are not separately specified. Two definitions may be distinguished: absolute inaccuracy and relative inaccuracy. Absolute inaccuracy is expressed in terms of units of the measuring quantity concerned, or as a fraction of the full-scale value. Relative inaccuracy relates the error to the actual measuring value.

Example 1.3

The data sheet of a volt meter with a four digit indicator and a full-scale value of 1.999 V specifies the instrument inaccuracy as ±0.05% FS ±0.1% of the indication $\pm\frac{1}{2}$ *digit.*

The absolute inaccuracy of a voltage of 1.036 V measured with this instrument equals: ±0.05 of 2 V (the approximate value of FS) plus ±0.1% of 1 V (approximate value of the indication) plus ±0.5 of 1 mV (the weight of the last digit), which amounts to ±2.5 mV in total.

The relative inaccuracy is the absolute inaccuracy divided by the indication so it is ±0.25%.

Inaccuracy is often confused with accuracy, the latter being complementary to it. When a specification list gives an accuracy of 1%, this hopefully means that there is an inaccuracy of 1% or an accuracy of 99%.

The sensitivity of a measuring system is defined as the ratio between a change in the output value and a change in the input value that causes that same output change. The sensitivity of a current-to-voltage converter is expressed in V/A, that of a linear position sensor in, for instance, mV/µm and that of an oscilloscope in, for instance, cm/V.

A measuring system is usually also sensitive to changes in quantities other than the intended input quantity, such as the ambient temperature or the supply voltage. These unwelcome sensitivities should be specified as well when this is necessary for a proper interpretation of the measurement result. To gain better insight into the effect of such false sensitivity it will be related to the sensitivity to the measurement quantity itself.

Example 1.4

A displacement sensor with voltage output has a sensitivity of 10 mV/mm. Its temperature sensitivity is -0.1 mV/K. Since -0.1 mV corresponds with a displacement of -10 μm, the temperature sensitivity can also be expressed as –10 μm/K. A temperature rise of 5°C will result in an apparent displacement of –50 μm.

Example 1.5

The sensitivity of a temperature sensor including the signal-conditioning unit is 100 mV/K. The signal conditioning part itself is also sensitive to (ambient) temperature and it appears to create an extra output voltage of 0.5 mV for each °C rise in ambient temperature (not necessarily the sensor temperature). The undesired temperature sensitivity is thus 0.5 mV/K or 0.5/100 = 5 mK/K. A change in ambient temperature of ± 10°C gives an apparent change in sensor temperature that is equal to ±50 mK.

Mathematically, the sensitivity is expressed as $S = dy/dx$. If output y is a linear function of input x then the sensitivity does not depend on x. In the case of a non-linear transfer function $y = f(x)$, S will depend on the input or output value (Figure 1.7). Users of measuring instruments prefer a linear response, because then the sensitivity can be expressed in terms of a single parameter and the output will not show harmonic distortion. The transfer of a system with slight non-linearity may be approximated by a straight line. The user should still know the deviation from the actual transfer as specified by the non-linearity.

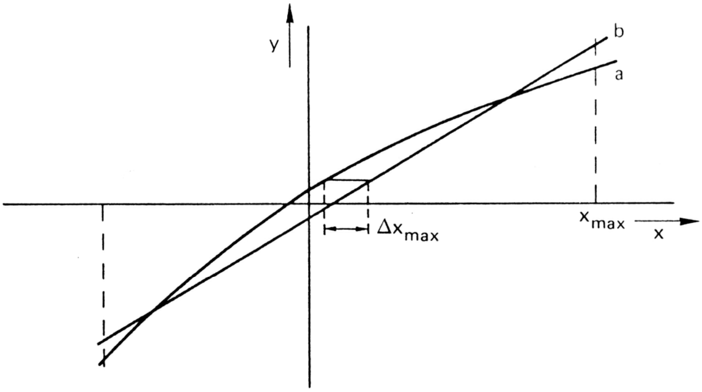


Figure 1.7. Example of a non-linear transfer characteristic, (a) real transfer, (b) linear approximation.

The non-linearity of a system is the maximum deviation in the actual transfer characteristic from a pre-described straight line. Manufacturers specify non-linearity in various ways, for instance, as the deviation in input or output units: $\Delta x_{\max}$ or $\Delta y_{\max}$, or as a fraction of FS: $\Delta x_{\max}/x_{\max}$. They may use different settings for the straight line: by passing through the end points of the characteristic, by taking the tangent through the point $x = 0$, or by using the best-fit (least-squares) line, to mention but a few possibilities.

Figure 1.8 depicts some particular types of non-linearity found in measuring systems: saturation, clipping and dead zone (sometimes also called cross-over distortion).

These are examples of static non-linearity, appearing even when inputs change slowly. Figure 1.9 shows another type of non-linearity, known as slew rate limitation, which only occurs when the input values change relatively fast. The output which is unable to keep up with the quickly changing input thus results in distortion at the output point. Slew rate is specified as the maximum rate of change in the output of the system.

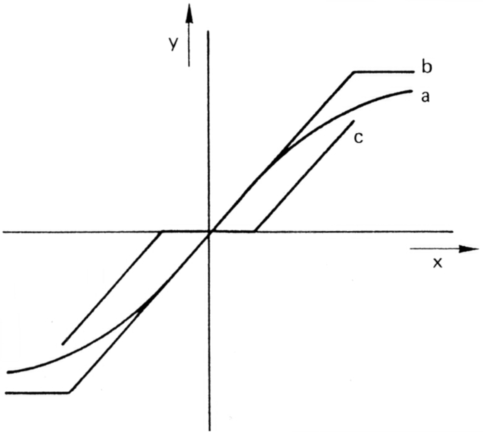


Figure 1.8. Some types of static non-linearity: (a) saturation, (b) clipping. (c) dead zone.

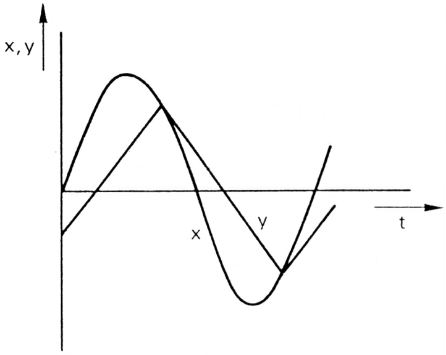


Figure 1.9. The effect of slew rate limitation on the output signal *y* at a sinusoidal input *x*.

Most measurement systems are designed in such a way that output is zero when input is zero. If the transfer characteristic does not intersect the origin ($x = 0, y = 0$) the system is said to have offset. Offset is expressed in terms of the input or the output quantity. It is preferable to specify the input offset so that comparisons with the real input quantity can be made. Non-zero offset arises mainly from component tolerances. Most electronic systems make it possible to compensate for the offset, either through manual adjustment or by means of manually or automatically controlled zero-setting facilities. Once adjusted to zero, the offset may still change due to temperature variations, changes in the supply voltage or the effects of ageing. This relatively slow change in the offset is what we call zero drift. It is the temperature-induced drift (the temperature coefficient or t.c of the offset) that is a particularly important item in the specification list.

Example 1.6

A data book on instrumentation amplifiers contains the following specifications for a particular type of amplifier:

input offset voltage:	*max. ±0.4 mV, adjustable to 0*
t.c. of the input offset:	*max.±6 μV/K*
supply voltage coeff.:	*40 μV/V*
long-term stability:	*3 μV/month*

There are two ways to determine the offset of any system. The first method is based on setting the output signal at zero by adjusting the input value. The input value for which the output is zero is the negative value of the input offset. The second method involves measuring the output at zero input value. When the output is still within the allowed range, the input offset simply becomes the measured output divided by the sensitivity.

Sometimes a system is deliberately designed with offset. Many industrial transducers have a current output that ranges from 4 to 20 mA (see Example 1.1). This facilitates the detection of cable fractures or a short-circuit so that such a defect is clearly distinguishable from a zero input.

The sensitivity of an electronic system may be increased to almost unlimited levels. There is, however, a limit to the usefulness of doing this. If one increases the sensitivity of the system its output offset will grow as well, to the limits of the output range. Even at zero input voltage, an ever-increasing sensitivity will be of no use, due to system noise interference. Electrical noise amounts to a collection of spontaneous fluctuations in the currents and voltages present in any electronic system, all of which arises from the thermal motion of the electrons and from the quantized nature of electric charge. Electrical noise is also specified in terms of input quantity so that its effect can be seen relative to that of the actual input signal.

The sensitivity of a system depends on the frequency of the signal to be processed. A measure of the useful frequency range is the frequency band. The upper and lower limits of the frequency band are defined as those frequencies where the power transfer has dropped to half its nominal value. For voltage or current transfer the criterion is $\frac{1}{2}\sqrt{2}$ of the respective nominal voltage and current transfer (Figure 1.10). The lower limit of the frequency band may be zero; the upper limit always has a finite value. The extent of the frequency band is called the bandwidth of the system expressed in Hz.

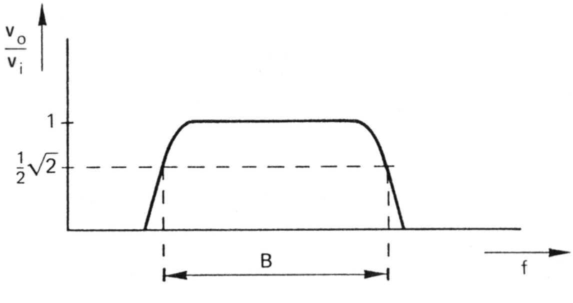


Figure 1.10. A voltage transfer characteristic showing the boundaries of the frequency band. The nominal transfer is 1, its bandwidth is B.

A frequent problem in instrumentation is the problem of how to determine the difference between two almost equal measurement values. Such situations occur when, for instance, big noise or interference signals are superimposed on relatively weak measurement signals. A special amplifier has been developed for these kinds of measurement problems, it is known as the differential amplifier (Figure 1.11). Such an amplifier, which is usually a voltage amplifier, has two inputs and one output. Ideally the amplifier is not sensitive to equal signals on both inputs (common mode signal), only to a difference between the two input signals (differential mode signals). In practice any differential amplifier will exhibit a non-zero transfer for common mode signals. A quality measure that relates to this property is the common mode rejection ratio or CMRR, which is defined as the ratio between the transfer for differential mode signals, v_o/v_d and common mode signals v_o/v_c. In other words, the CMRR is the ratio of a common mode input signal and a differential mode input signal, both of which give equal output. An ideal differential amplifier has a CMRR, which is infinite.

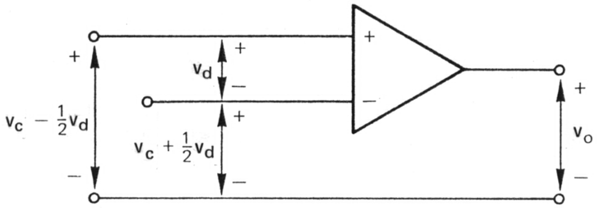


Figure 1.11. An ideal differential amplifier is insensitive to common mode signals (v_c) and amplifies only the differential signal v_d.

Example 1.7

A system with a CMRR of 10^5 *is used to determine the difference between two voltages, both about 10 V high. The difference appears to be 5 mV. The inaccuracy of this result, due to the finite CMRR, is ±2% because the common mode voltage produces an output voltage that is equivalent to that of a differential input voltage of* $10/10^5 = 0.1$ *mV.*

The final system property to be discussed in this chapter has to do with reliability. There is always a chance that a system will fail after a certain period of time. Such properties should be described according to probability parameters, one of these parameters being the reliability $R(t)$ of the system. This is defined as the probability that the system will function correctly (in accordance with its specifications) up to the time t (provided that the system has operated within the permitted range). It should be clear that R diminishes as time elapses so that the system becomes increasingly less reliable.

The system parameter R has the disadvantage that it changes over the course of time. Better parameters are the mean-time-to-failure (MTTF) and the failure rate $\lambda(t)$. The MTTF is the mean time that passes up until the moment when the system fails; it is its mean lifetime.

Example 1.8

An incandescent lamp is guaranteed for 1000 burning hours. This means that lamps from the series to which this lamp belongs will burn, on average, for 1000 hours. Some lamps may fail earlier or even much earlier while others may burn longer.

The failure rate $\lambda(t)$ is defined as the fraction of failing systems per unit of time relative to the total number of systems functioning properly at time t. The failure rate appears to be constant during a large part of the system's lifetime. If the failure rate is constant in terms of time, it is equal to the inverse of the MTTF.

> ***Example 1.9***
>
> *Suppose an electronic component has an MTTF equal to 10^5 hours. Its failure rate is the inverse, 10^{-5} per hour or 0.024% per day or 0.7% per month. Thus, if one takes a certain collection of correctly functioning components 0.024% will fail daily.*

The failure rate of electronic components is extremely low when used under normal conditions. For example, the failure rate of metal film resistors with respect to an open connection is approximately 5×10^{-9} per hour. The reliability of many electronic components is well known. However, it is very difficult to determine the reliability of a complete electronic measurement system from the failure rates of the individual components. This is a reason why the reliability of complex systems is seldom specified.

SUMMARY

System functions

- The three main functions of an electronic measurement system are
  - data acquisition
  - data processing
  - data distribution
- The conversion of information of a physical quantity into an electrical signal is known as transduction. Transduction is carried out with an input transducer or sensor. The inverse process is carried out with an output transducer or actuator
- The main operations completed with analog measurement signals are: amplification, filtering, modulation, demodulation and analog-to-digital conversion.
- AD conversion comprises three elements: sampling, quantization and coding.
- Multiplexing is a technique that facilitates the simultaneous transport of various signals through a single channel. There are two different possible ways of doing this: by time multiplexing and by frequency multiplexing. The inverse process is called demultiplexing.

System specifications

- The main specifications of any measurement system are: operating range (including measuring range), resolution, accuracy, inaccuracy, sensitivity, non-linearity, offset, drift and reliability.
- Some possible types of non-linearity are: saturation, clipping, dead zone, hysteresis and slew rate limitation.
- The bandwidth of a system is the frequency span between frequencies where the power transfer has dropped to half the nominal value or where the voltage or current transfer has dropped to $\frac{1}{2}\sqrt{2}$ of the nominal value.

- The common-mode rejection ratio is the ratio between the transfer of differential mode signals and common mode signals, or: the ratio between a common mode input and a differential mode input, both producing equal outputs.
- Noise is the phenomenon of spontaneous voltage or current fluctuations occurring in any electronic system. It fundamentally limits the accuracy of a measurement system.
- The reliability of a system can be specified in terms of the reliability $R(t)$, the failure rate $\lambda(t)$ and the mean-time-to-failure MTTF. For systems with constant failure rate, $\lambda = 1/\text{MTTF}$.

EXERCISES

System functions

1.1 What is meant by multiplexing? Describe the process of time multiplexing.

1.2 Discuss the difference between the requirements for a multiplexer used for digital signals and one used for analog signals.

1.3 Compare an AD converter in a centralized system with that of a distributed system from the point of view of the conversion time.

System specifications

1.4 What would be the reason for putting a factor $1/\sqrt{2}$ in the definition of the bandwidth for voltage transfer, instead of a factor $\frac{1}{2}$?

1.5 What is a differential voltage amplifier? What is meant by the CMRR of such an amplifier?

1.6 The CMRR of a differential voltage amplifier is specified as CMRR > 10^3, its voltage gain is $G = 50$. The two input voltages have values $V_1 = 10.3$ V, $V_2 = 10.1$ V. What is the possible output voltage range?

1.7 The slew rate of a voltage amplifier is 10 V/μs, its gain is 100. The input is a sinusoidal voltage with amplitude A and frequency f.
a. Suppose $A = 100$ mV, what would be the upper limit of the frequency where the output would show no distortion?
b. Suppose $f = 1$ MHz; up to what amplitude can the input signal be amplified without introducing distortion?

1.8 A voltage amplifier is specified as follows: input offset voltage at 20°C is < 0.5 mV, the temperature coefficient of the offset is < 5 μV/K. Calculate the maximum input offset that might occur within a temperature range of 0 to 80 °C.

1.9 The relation between the input quantity x and the output quantity y of a system is given as: $y = \alpha x + \beta x^2$, with $\alpha = 10$ and $\beta = 0.2$. Find the non-linearity relative to the line $y = \alpha x$, for the input range $-10 < x < 10$.

2 Signals

Physical quantities that contain detectable messages are known as signals. The information carrier in any electrical signal is a voltage, a current, a charge or some other kind of electric parameter.

The message contained in such a signal may constitute the result of a measurement but it can also be an instruction or a location code (like, for instance, the address of a memory location). The nature of the message cannot be deduced from its appearance. The processing techniques for electronic signals are as they are, regardless of the contents or nature of the message.

The first part of this chapter will concentrate on the characterization of signals and the various values of signals in terms of time functions. Signals may alternatively be characterized according to their frequency spectrum. In the case of periodic signals, the frequency spectrum is determined by means of Fourier expansion.

The second part of this chapter deals with aperiodic signals, in particular: noise, stochastic and sampled signals.

2.1 Periodic signals

2.1.1 A classification of signals

There are many ways to classify signals but one way is on the basis of their dynamic properties.

- Static or DC signals (DC = direct current, a term that is also applied to voltages): the signal value remains constant during the measuring time interval.
- Quasi-static signals: the signal value varies just a little, according to a given physical quantity. An example of a quasi-static signal is drift.
- Dynamic signals: the signal value varies significantly during the observation period. Such signals are also termed AC signals (AC = alternating current or alternating voltages).

Another way to distinguish signals is on the basis of the difference between deterministic and stochastic signals. What characterizes a stochastic signal is the fact that its exact value is impossible to predict. Most measurement signals and interference signals, such as noise, belong to this category. Examples of deterministic signals are:

- Periodic signals, characterized as $x(t) = x(t + nT)$, in which T is the time of a signal period and n the integer.
- Transients, like the response of a system to a pulse-shaped input: the signal can be repeated (in other words predicted) by repeating the experiment under the same conditions.

A third possibility is to consider continuous and discrete signals. The continuity may refer both to the time scale and to the amplitude scale (the signal value). Figure 2.1 shows the four possible combinations. Figure 2.1b represents a sampled signal and Figure 2.1c illustrates a quantized signal, as mentioned in Chapter 1. A quantized signal that only has two levels is called a binary signal.

Finally, we shall contemplate the distinction between analog and digital signals. As with many technical terms (especially electronic terms) the meaning here becomes rather fuzzy. In ordinary terms, digital signals are sampled, time-discrete and binary-coded, as in digital processors. Analog signals refer to time-continuous signals that have a continuous or quantized amplitude.

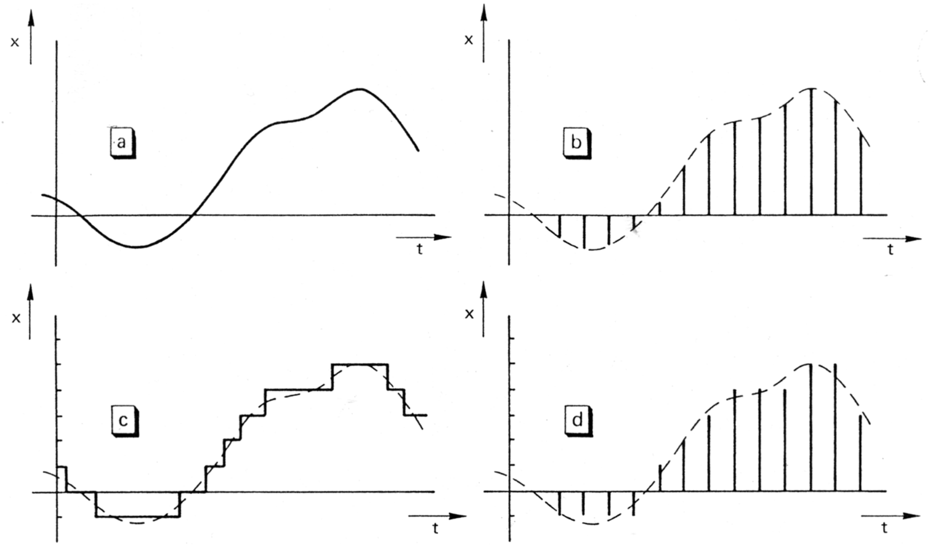


Figure 2.1. Continuous and discrete signals: (a) continuous in time and in amplitude, (b) time discrete, continuous amplitude (sampled signal), (c) discrete amplitude, continuous in time (quantized signal), (d) discrete both in time and amplitude.

2.1.2 Signal values

Amplitude-time diagrams of the type given in Figure 2.1, which represent the signal value for each moment within the observation interval, are the most complete kinds of signal descriptions. Invariably it is not necessary to give that much information about the signal; a mere indication of a particular signal property would suffice. Some such simple characteristic signal parameters are listed below. The parameters are valid for an observation interval $0 < t < \tau$.

peak value: $x_p = \max\{|x(t)|\}$

peak-to-peak value: $x_{pp} = \max\{x(t)\} - \min\{x(t)\}$

mean value: $x_m = \frac{1}{\tau}\int_0^\tau x(t)dt$

mean absolute value: $|x|_m = \frac{1}{\tau}\int_0^\tau |x(t)|dt$

root-mean-square value: $x_{eff} = \sqrt{\frac{1}{\tau}\int_0^\tau x^2(t)dt}$

mean signal power: $P_m = \frac{1}{\tau}\int_0^\tau x^2(t)dt$

The peak and peak-to-peak values are important in relation to the limits of the signal range of an instrument. The mean value is used when it is only the DC or quasi-DC component of a signal that counts. The rms value is a parameter related to the signal power content. An arbitrarily shaped AC current with an rms value of I (A) which flows through a resistor will produce just as much heat as a DC current with a (DC) value of I (A). Note that the rms value is the square root of the mean power.

Example 2.1
The mathematical description of a sinusoidal signal is:

$$x(t) = A\sin\omega t = A\sin\frac{2\pi}{T}t$$

where A is the amplitude, $f = \omega/2\pi$ the frequency and $T = 1/f$ the period time. Figure 2.2 shows one period of this signal while illustrating the characteristic parameters defined above. If these definitions are applied to the sine wave this will result in the following values:

$$x_p = A$$

$$x_{pp} = 2A$$

$$x_m = 0$$

$$|x|_m = \frac{2A}{\pi}$$

$$x_{rms} = \tfrac{1}{2}A\sqrt{2}$$

As the shapes of all periods are equal these values also apply to a full periodical sine wave.

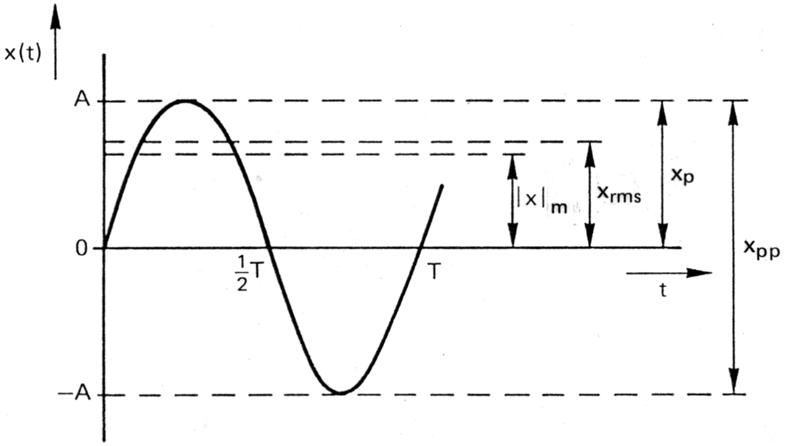


Figure 2.2. Signal values for a sine wave signal.

Many rms voltmeters do not actually measure the rms value of the input signal but rather the mean of the absolute value, $|x|_m$, which can be realized with the aid of a very simple electronic circuit. The two values are not, however, the same. To obtain an rms indication such instruments have to be calibrated in terms of rms. Since both signal parameters depend on the signal shape the calibration will only be valid for the particular signal used while calibrating. Generally, rms meters are calibrated for sinusoidal inputs. Example 2.1 shows that the mean absolute value should be multiplied (internally) by $\frac{1}{4}\pi\sqrt{2}$, about 1.11, to obtain the rms value. Such instruments only indicate the proper rms value for sine shaped signals.

Some voltmeters indicate the "true rms" value of the input voltage. A true rms meter functions differently from those described above. Some of them use a thermal converter to directly obtain the rms value of the input signal. The indication is true for almost all types of input voltages.

2.1.3 Signal spectra

Any periodic signal can be divided into a series of sinusoidal sub-signals. If the time of one period is T, then the frequencies of all the sub-signals will be multiples of $1/T$. There are no components with other frequencies. The lowest frequency which is equal to $1/T$ is known as the fundamental frequency of the signal.

The subdividing of a periodic signal into its sinusoidal components is known as “Fourier expansion of the signal”. The resultant series of sinusoids is thus a Fourier series. Fourier expansion can be described mathematically as follows:

$$\begin{aligned} x(t) &= a_0 + a_1\cos\omega_0 t + a_2\cos 2\omega_0 t + a_3\cos 3\omega_0 t + \cdots \\ &\quad + b_1\sin\omega_0 t + b_2\sin 2\omega_0 t + b_3\sin 3\omega_0 t + \cdots \\ &= a_0 + \sum_{n=1}^{\infty}\left(a_n\cos n\omega_0 t + b_n\sin n\omega_0 t\right) \qquad (2.1) \\ &= a_0 + \sum_{n=1}^{\infty} c_n\cos\left(n\omega_0 t + \varphi_n\right) \end{aligned}$$

These three representations are identical; the second is an abbreviated form of the first. In the third representation the corresponding sine and cosine terms are combined in a single cosine with the new amplitude c and the phase angle φ, which satisfies the relations:

$$c_n = \sqrt{a_n^2 + b_n^2} \qquad \varphi_n = \arctan\left(b_n/a_n\right) \qquad (2.2)$$

The coefficients a_n, b_n and c_n are the Fourier coefficients of the signal. Each periodic signal can be written as a collection of sinusoidal signals with amplitudes given by the Fourier coefficients and frequencies that are multiples of the fundamental signal frequency.

The term a_0 in Equation (2.1) is nothing other than the mean value of the signal $x(t)$: the mean value must be equal to that of the complete series, and the mean of each sine signal is zero. All sine and cosine terms of the Fourier series have a frequency that is a multiple of the fundamental, f_0; they are termed the harmonic components or the signal (i.e. if the signal were made audible by a loudspeaker a perfect "harmonic" sound would be heard). The component with a frequency of $2f_0$ is the second harmonic, $3f_0$ is the third harmonic, and so on.

The shape of a periodic signal is reflected in its Fourier coefficients. We can illustrate the Fourier coefficients as a function of the corresponding frequency. Such a diagram is called the frequency spectrum of the signal (Figure 2.3). Usually the amplitude of the combined sine and cosine terms is plotted so that the coefficient is c_n as in Equation (2.1).

The Fourier coefficients are related to the signal shape. They can be calculated using the transformation formulas given in Equations (2.3):

$$\begin{aligned} a_0 &= \frac{1}{T}\int_{t_0}^{t_0+T} x(t)dt \\ a_n &= \frac{2}{T}\int_{t_0}^{t_0+T} x(t)\cos n\omega_0 t dt \qquad (2.3) \\ b_n &= \frac{2}{T}\int_{t_0}^{t_0+T} x(t)\sin n\omega_0 t dt \end{aligned}$$

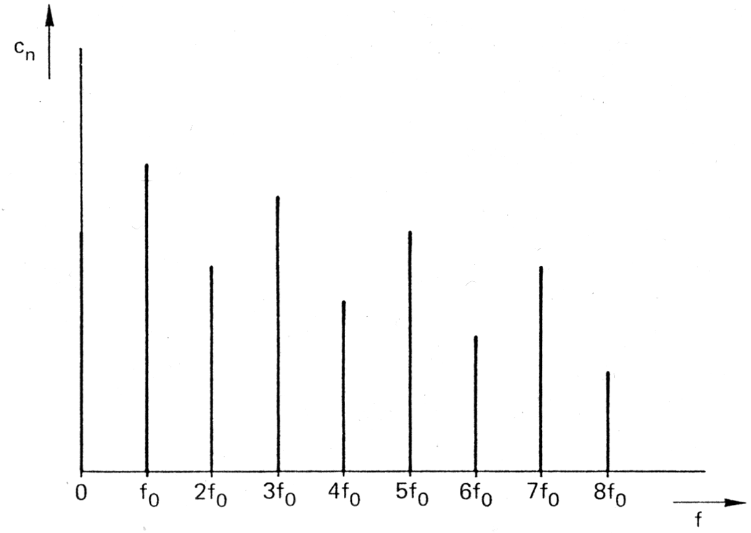


Figure 2.3. An example of a frequency spectrum of a periodic signal.

These equations present the discrete Fourier transform for real coefficients. In general, the Fourier series has an infinite length. The full description of a signal, according to its spectrum, requires an infinite number of parameters. Fortunately, the coefficients tend to diminish when frequencies increase. One remarkable property of the coefficients is that the first N elements of the series constitute the best approximation of the signal in N parameters.

Example 2.2

The Fourier coefficients of the square-shaped signal given in Figure 2.4a, calculated with Equations (2.2), are:

$$a_0 = 0$$

$$a_n = 0$$

$$b_n = \frac{2A}{n\pi}(1 - \cos n\pi)$$

Apparently, its Fourier series is described as:

$$x_1(t) = \frac{4A}{\pi}\left(\sin\omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \frac{1}{5}\sin 5\omega_0 t + \cdots\right)$$

The signal appears to be composed only of sinusoids with frequencies that are odd multiples of the fundamental frequency.

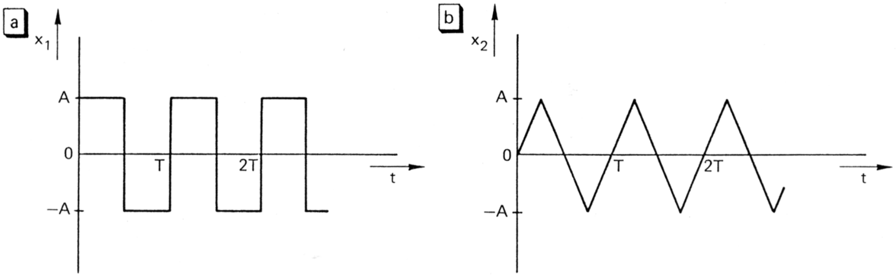


Figure 2.4. Examples of two periodical signals: (a) a square wave signal, (b) a triangular signal.

Example 2.3
Using the same transformation formulas, the frequency spectrum of the triangular signal in Figure 2.4b is calculated as:

$$x_2(t)=\frac{8A}{\pi^2}\left(\sin\omega_0 t-\frac{1}{9}\sin 3\omega_0 t+\frac{1}{25}\sin 5\omega_0 t+\cdots\right)$$

and consists of components that have exactly the same frequencies, but different amplitudes.

According to the theory of Fourier, any periodic signal can be split up into sinusoidal components with discrete frequencies. The signal in question has a discrete frequency spectrum or a line spectrum. Obviously, one can also create an arbitrary periodic signal by adding the required sinusoidal signals with the proper frequencies and amplitudes. This particular composition of periodic signals is used in synthesizers.

The Fourier transform is also applicable to aperiodic signals. It appears that such signals have a continuous frequency spectrum. A continuous spectrum does not have any individual components, but the signal is expressed in terms of amplitude density rather than amplitude. A more usual way of presenting a signal is according to its power spectrum, that is, its spectral power (W/Hz) as a function of frequency.

Figure 2.5 shows the power spectra of two different signals. One signal varies gradually over the course of time while the other is much faster. One can imagine the first signal being composed of sinusoidal signals with relatively low frequencies. Signal (b) contains components with frequencies that are higher. This is clearly illustrated in the corresponding frequency spectra of the signals: the spectrum of signal (a) covers a small range of frequency and its bandwidth is low. Signal (b) has a much wider bandwidth.

The relationship between the signal shape (time domain) and its spectrum (frequency domain) is also illustrated in Figure 2.6 which shows the spectrum of two periodic signals, one with very sharp edges (the rectangular signal) and another that does not vary so quickly (a rectified sine wave). Clearly the high frequency components of the rectangular wave are much larger than those of the clipped sine wave.

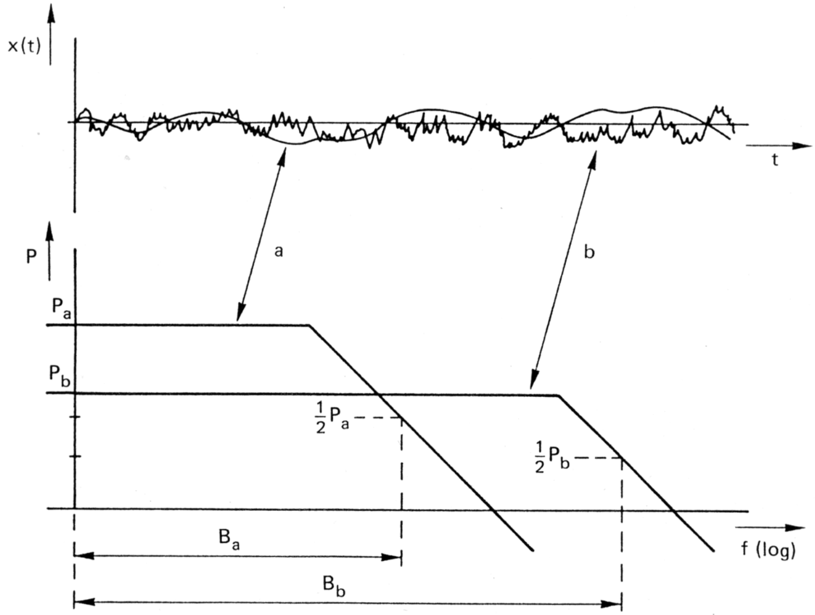


Figure 2.5. The amplitude-time diagram of two signals a and b, and the corresponding power spectra. Signal a varies slowly, and has a narrow bandwidth. Signal b moves quickly; it has a larger bandwidth.

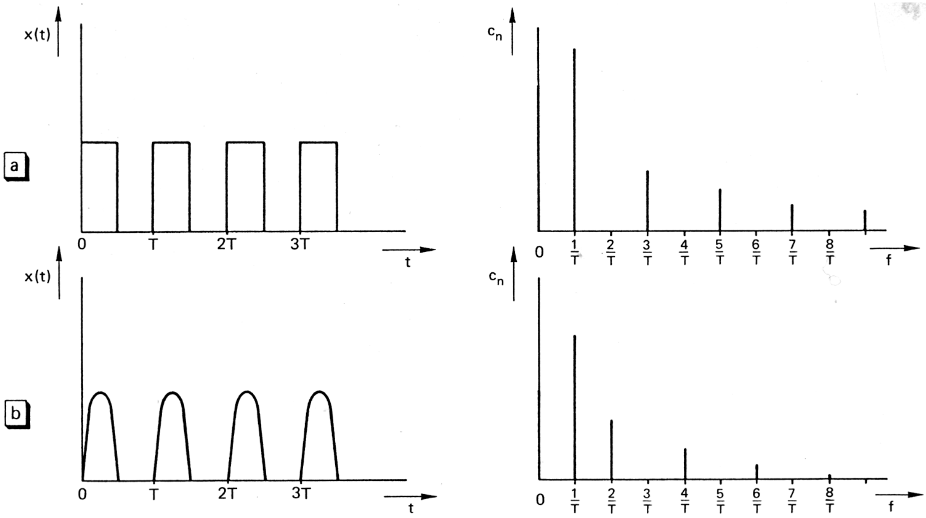


Figure 2.6. The amplitude-time diagram and the frequency spectrum of (a) a rectangular signal, (b) the positive half of a sine wave.

The bandwidth of a signal is defined in a similar way to that for systems. The bandwidth is the part of the signal spectrum found between the frequencies where the power spectrum has dropped to half of its nominal or maximal value. In the case of an amplitude spectrum the boundaries are defined at 1/√2 of the nominal amplitude density.

A measurement system can only cope with signals that have a bandwidth up to that of the system itself. Signals with high frequency components require a wideband processing system. The bandwidth of the measuring instrument should correspond to that of the signals being processed.

Randomly varying signals or noise also have continuous frequency spectra. Some types of noise (in particular thermal induced electrical noise) have constant spectral power P_n (W/Hz) where, up to a certain maximum frequency, the power spectrum is flat. Like white light, such signals are called white noise and contain equal wavelength components (colors) within the visible range. Noise can also be specified as spectral voltage or spectral current, expressed respectively in V/√Hz and A/√Hz.

Example 2.4
Let the respective spectral power, spectral voltage and spectral current density of white noise be P_n W/Hz, V_n V/√Hz and I_nA/√Hz. The noise power, noise voltage and noise current of this signal, measured within a frequency band of 200 to 300 Hz amount to: $100 \cdot P_n$ W, $10 \cdot V_n$ V and $10 \cdot I_n$ A.

2.2 Aperiodic signals

In this section we shall extend the Fourier expansion definition to include aperiodic signals and use the result to deduce the spectrum of sampled signals. Stochastic signals (whether they be continuous or discrete) can be described in three ways: according to their time domain properties (e.g. time average, rms value), their frequency domain properties (amplitude spectrum, power spectrum) or their amplitude properties (expressing the signal value with probability parameters).

2.2.1 Complex Fourier series

In the first part of this chapter we showed how the Fourier expansion of a periodic signal can lead to a series of (real) sine and cosine functions. The complex Fourier expansion was established using Euler's relation

$$e^{\pm jz} = \cos z \pm j \sin z \tag{2.4}$$

Solving sin z and cos z, and replacing the real goniometric functions in (2.1) with their complex counterparts we obtain:

$$x(t) = a_0 + \sum_{n=1}^{\infty} \left[\frac{a_n}{2}\left(e^{jn\omega t} + e^{-jn\omega t}\right) + \frac{b_n}{2j}\left(e^{jn\omega t} - e^{-jn\omega t}\right) \right] \tag{2.5}$$

Using the substitutions $C_0 = a_0$, $C_n = \frac{1}{2}(a_n - jb_n)$ and $C_{-n} = \frac{1}{2}(a_n + jb_n)$ this can be simplified to

$$x(t) = C_0 + \sum_{n=1}^{\infty}\left(C_n e^{jn\omega t} + C_{-n}e^{-jn\omega t}\right) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega t} \tag{2.6}$$

What this results in is the complex Fourier series. Similarly, the complex form of Equations (2.3) becomes:

$$C_n = \frac{1}{T}\int_{t_0}^{t_0+T} x(t)e^{-jn\omega t}dt \qquad n = 0,\ 1,\ 2,\ ... \tag{2.7}$$

the discrete complex Fourier transform. The complex Fourier coefficients C_n can easily be derived from the real coefficients using the relations

$$\begin{aligned} |C_n| &= \tfrac{1}{2}\sqrt{a_n^2 + b_n^2} \qquad n \neq 0 \\ \arg C_n &= \arctan\frac{-b_n}{a_n} \end{aligned} \tag{2.8}$$

As C_n is complex, the complex signal spectrum consists of two parts: the amplitude spectrum – a plot of $|C_n|$ versus frequency and the phase spectrum, a plot of arg C_n versus frequency.

Example 2.5

The complex Fourier series of the rectangular signal in Figure 2.4a is calculated as follows: $C_0 = 0$, so $|C_0| = 0$ and $\arg C_0 = 0$. As $C_n = \frac{1}{2}(a_n - jb_n)$, its modulus and argument are:

$$|C_n| = \tfrac{1}{2}\sqrt{a_n^2 + b_n^2} = \frac{A}{n\pi}(1 - \cos n\pi) \qquad n = 1,2,...$$

and

$$\arg C_n = \arctan\frac{-b_n}{a_n} = -\frac{\pi}{2} \qquad n = 1,2,...$$

The amplitude and phase spectra are depicted in Figure 2.7.

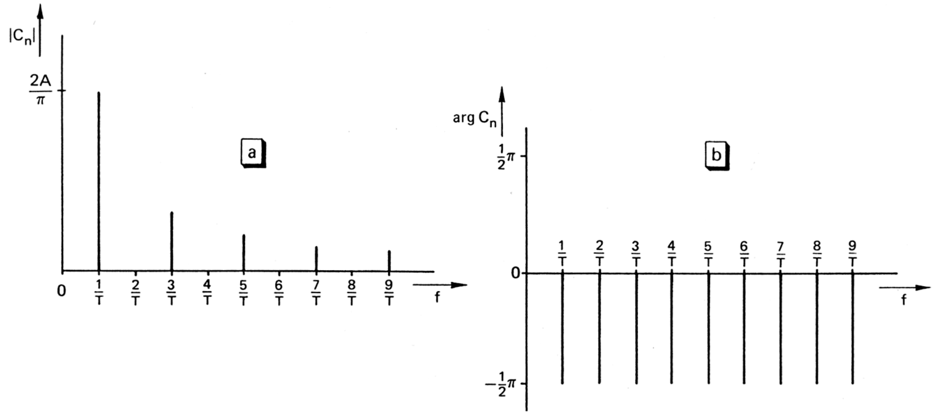


Figure 2.7. (a) amplitude spectrum and (b) phase spectrum of the rectangular signal from Figure 2.4a.

2.2.2 The Fourier integral and the Fourier transform

To obtain the Fourier expansion of a non-periodic signal we start with the discrete complex Fourier series for periodic signals as given in Equations (2.6) and (2.7). Consider one period of this signal. Replace t_0 with $-\frac{1}{2}T$ and let T approach infinity. Then:

$$x(t)=\lim_{T\to\infty}\sum_{n=-\infty}^{\infty}C_n e^{jn\omega t}=\lim_{T\to\infty}\sum_{n=-\infty}^{\infty}e^{jn\omega t}\left(\frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T}x(t)e^{-j\omega nt}dt\right)\tag{2.9}$$

When taking the limit for T→∞, the summation becomes an integration, $n\omega$ changes to ω and $T = (1/2\pi)\omega$ becomes $(1/2\pi)d\omega$:

$$x(t)=\int_{-\infty}^{\infty}e^{j\omega t}\left(\int_{-\infty}^{\infty}x(t)e^{-j\omega t}dt\right)\frac{d\omega}{2\pi}\tag{2.10}$$

With

$$X(\omega)=\int_{-\infty}^{\infty}x(t)e^{-j\omega t}dt\tag{2.10a}$$

this results in:

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)e^{j\omega t}d\omega\tag{2.10b}$$

$X(\omega)$ is the complex Fourier transform of $x(t)$. Both $x(t)$ and $X(\omega)$ give a full description of the signal, the first in the time domain, the other in the frequency domain. Equations (2.10) transform the signal from the time domain to the frequency domain and vice versa. The modulus and argument provided by $X(\omega)$ describe the frequency spectrum of $x(t)$. In general, this is a continuous spectrum, extending from $-\infty$ to $+\infty$ also containing (in a mathematical sense) negative frequencies.

To find the Fourier transform of the product of two signals $x_1(t)$ and $x_2(t)$, we first define a particular function, the convolution integral:

$$g(\tau) = \int_{-\infty}^{\infty} x_1(t) x_2(\tau - t) dt \tag{2.11}$$

This is the product of $x_1(t)$ and the shifted and back-folded function $x_2(t)$ (Figure 2.8) integrated over an infinite time interval. The convolution function $g(\tau)$ is also denoted as:

$$g(\tau) = x_1(t) * x_2(t) \tag{2.12}$$

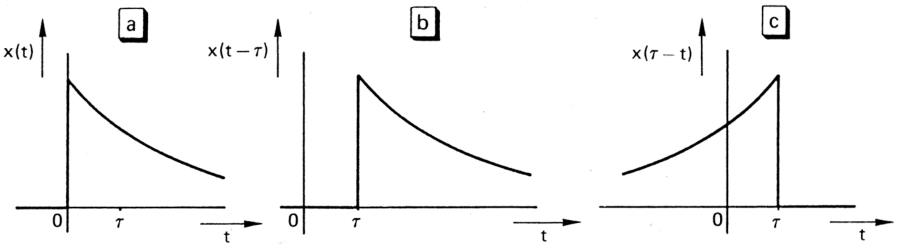


Figure 2.8. (a) original function, (b) shifted over $t = \tau$; (c) shifted and back-folded.

The Fourier transform of $g(\tau)$ is:

$$\begin{aligned} F\{g(\tau)\} &= \int_{-\infty}^{\infty} g(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x_1(t) x_2(\tau - t) dt \right) e^{-j\omega\tau} d\tau \end{aligned} \tag{2.13}$$

Splitting up the term $e^{-j\omega\tau}$ into $e^{-j\omega t} \cdot e^{-j\omega(\tau - t)}$ and changing the order of integration results in:

$$\begin{aligned} F\{g(\tau)\} &= \int_{-\infty}^{\infty} x_1(t) e^{-j\omega t} \left(\int_{-\infty}^{\infty} x_2(\tau - t) e^{-j\omega(\tau - t)} d\tau \right) dt \\ &= X_1(\omega) \cdot X_2(\omega) \end{aligned} \tag{2.14}$$

The Fourier transform of two convoluted functions $x_1(t)$ and $x_2(t)$ therefore equals the product of the individual Fourier transforms. Similarly, the Fourier transform of the convolution $X_1(\omega)*X_2(\omega)$ equals $x_1(t).x_2(t)$.

The Fourier transform is used to calculate the frequency spectrum of both deterministic and stochastic signals. The Fourier transform is only applicable to functions that satisfy the following inequality:

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \tag{2.15}$$

In order to calculate the frequency characteristics of functions that do not satisfy (2.17), another kind of transformation should be used.

2.2.3 A description of sampled signals

In this section we will calculate the spectrum of a sampled signal. We will consider sampling over equidistant time intervals. The sampling of a signal $x(t)$ can be seen as the multiplication of $x(t)$ by a periodic, pulse-shaped signal $s(t)$, as indicated in the left section of Figure 2.9. The sampling width is assumed to be zero.

As $y(t)$ is the product of $x(t)$ and $s(t)$, the spectrum of $y(t)$ is described by the convolution of the Fourier transforms which are $X(f)$ and $S(f)$ respectively. $S(f)$ is a line spectrum because $s(t)$ is periodical. The height of the spectral lines are all equal when the pulse width of $s(t)$ approaches zero (their heights decrease with frequency at finite pulse width). $X(f)$ has a limited bandwidth, its highest frequency being B.

The first step towards establishing $Y(f) = X(f)*S(f)$ is to back-fold $S(f)$ along a line $f = \xi$ in order to find the function $S(\xi - f)$ using ξ, a new frequency variable. As $S(f)$ is a symmetric function, $S(\xi - f)$ is found by simply moving $S(f)$ over a distance, ξ, along the f-axis. For each ξ, the product of the shifted version of $S(f)$ and $X(f)$ is then integrated over the full frequency range. This product only consists of a single line at $f = \xi$, as long as one pulse component of $S(\xi - f)$ falls within the band $X(f)$. It is only that line which contributes to the integral because for all the other values the product is zero. The convolution process results in periodically repeated frequency bands known as the alias of the original band (Figure 2.9c, right).

In the section above it is assumed that the bandwidth, B, of $x(t)$ is smaller than half the sampling frequency. In such cases the multiple bands do not overlap and the original signal can be completely reconstructed without any information loss. With a larger bandwidth of $x(t)$ or when the sampling frequency is below $2/B$, the multiple bands in the spectrum of the sampled signal will overlap, thus preventing signal reconstruction and loss of information (or signal distortion). The error derived from such overlapping is called aliasing error and it occurs when the sample frequency is too low. The criterion required to avoid such aliasing errors is a sampling frequency of at least twice the highest frequency component of the analog signal. This result is known as the Shannon sampling theorem and it gives the theoretical lower limit of the sampling rate. In practice, one would always choose a much higher sampling frequency so as to facilitate the reconstruction of the original signal.

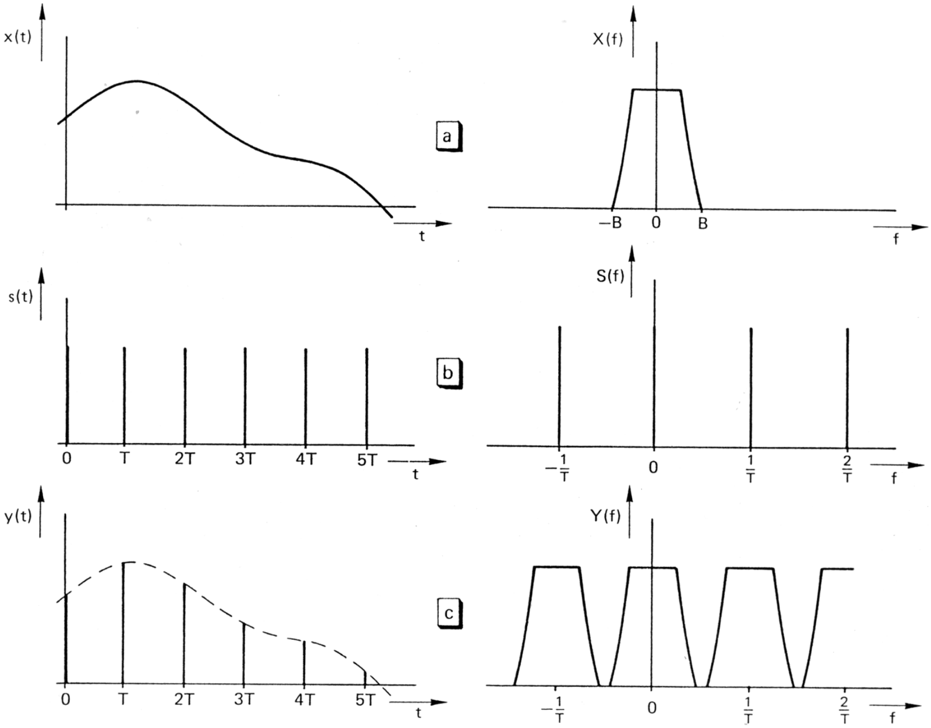


Figure 2.9. The amplitude-time diagram and the frequency spectrum of (a) an analogue signal $x(t)$, (b) a pulse shaped signal s(t) and (c) the product $y(t) = x(t) \cdot s(t)$.

2.2.4 A description of stochastic signals

In this section we will describe stochastic signals in the amplitude domain in terms of statistical parameters. At the end of the description, certain parameters will be related to the signal parameters in the time domain.

We may distinguish between continuous and discrete stochastic signals or variables which are denoted as x and $\underline{x}$, respectively. A discrete stochastic signal can result from converting a continuous stochastic signal into a digital signal using an AD converter. Again, a full description of a stochastic signal in the time domain requires a great deal of information. For most applications a rough description giving the statistical parameters will suffice.

Let us just consider a signal source with known statistical properties that generates a continuous, stochastic signal $x(t)$. Although it is impossible to precisely predict the signal we can estimate its value which will depend on the nature of the source or the process that generates the signal. For instance, we know the probability P that the signal value $x(t)$ at any given moment will not exceed a certain value x. This probability, which depends on the value x, is called the distribution function of $x(t)$ and is denoted as $F(x) = P\{x(t) < x\}$. Figure 2.10a gives an example of such a distribution function. From the definition it automatically follows that $F(x)$ is a monotonically non-decreasing function of x, that $F(x \to \infty) = 1$ and that $F(x \to -\infty) = 0$. Another important statistic parameter is the derivative of $F(x)$; the probability density (function): $p(x) = dF(x)/dx$. This function describes the probability of $x(t)$ having a value between x and $x + dx$ (Figure 2.10b). As the sum of the probability of all possible values is exactly 1, $p(x)$ satisfies:

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{2.16}$$

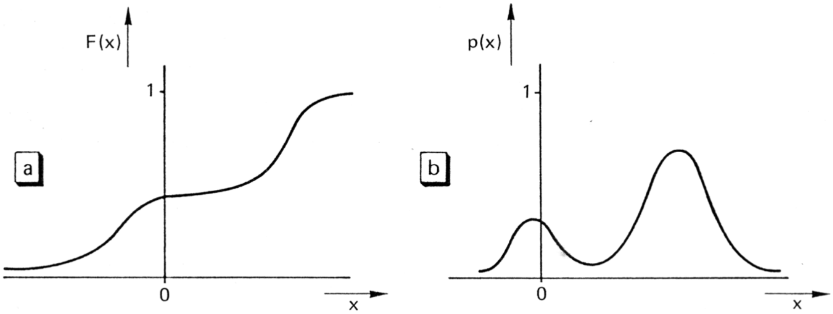


Figure 2.10. An example of (a) a distribution function and (b) the corresponding probability function.

If $p(x)$ is known, $F(x)$ can be found through:

$$\int_{-\infty}^{\infty} p(x)dx = F(x) - F(-\infty) = F(x) \tag{2.17}$$

Many physical processes are governed by a normal or Gaussian distribution function. The probability density of the produced signals is given as:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{2.18}$$

The meaning of the parameters μ and σ will be explained later in this section. The corresponding distribution function is:

$$F(x) = \int_{-\infty}^{x} p(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2} dx \tag{2.19}$$

For the normalized form, which is $\mu = 0$ and $\sigma = 1$, the numerical values of this integral function can be found in mathematical tables.

The distribution function $F(x)$ of a discrete stochastic variable $\underline{x}$ is the probability that $\underline{x}$ does not exceed value x, so $F(x) = P\{\underline{x} < x\}$.

> ***Example 2.6***
> *Suppose that an electrical voltage can only have two values: 0 V and 2 V, with a probability of 2/3 for the value of 0 V. Figure 2.11 gives the distribution function and the corresponding probability density function for this binary signal.*

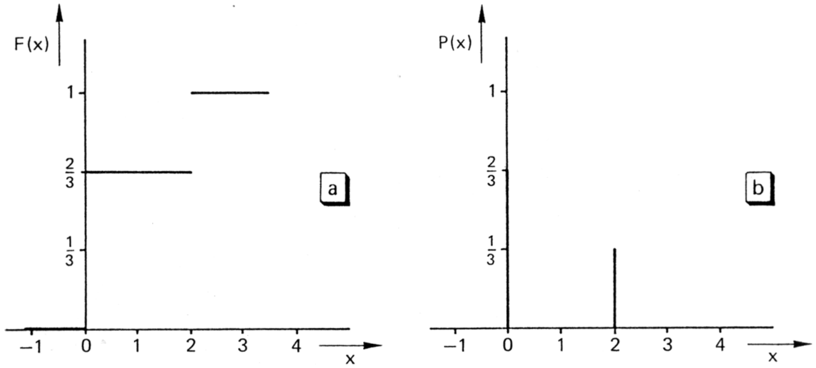


Figure 2.11. (a) The distribution function and (b) the probability density function of a binary signal of which $P(\underline{x}=0) = 2/3$.

Other statistical parameters used to characterize a continuous or discrete stochastic variable are the mean or average value, the expectancy or first order moment, the second order moment and the standard deviation. We will first discuss these parameters in connection with discrete variables.
The mean value of a discrete stochastic variable $\underline{x}$ is defined as:

$$\bar{x}_m = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2.20)$$

where N is the number of all x_i values considered in a certain experiment. The expectancy or the first moment for $\underline{x}$ is defined as:

$$E(\bar{x}) = \frac{1}{N}\sum_{i=1}^{N} p(x_i)x_i \qquad (2.21)$$

which can be seen as the weighted average of the all x_i values. Only with an infinite number of values ($N \to \infty$), will the (algebraic) mean x_m approach the expected value of $E(\underline{x})$. This can be explained as follows. The probability $p(x_k)$ of a certain output x_k is equal to N_k/N, where N_k is the number of x_k outputs. Suppose there are m different outputs. Then:

$$\begin{aligned}\lim_{N\to\infty} \bar{x}_m &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} N_k x_k \\ &= \sum_{k=1}^{m} p(x_k)x_k = \sum_{i=1}^{N} p(x_i)x_i = E(\bar{x})\end{aligned} \tag{2.22}$$

Example 2.7

The mean of the number of dots per throw after 100 throws have been made with a die is simply the sum of all the dots thrown divided by 100. The result is, for instance, 358/100 = 3.58.

As the probability for any result i = 1 to 6 is just 1/6, the expectancy is:

$$E(\bar{x}) = \sum_{i=1}^{6} \frac{1}{6} i = \frac{21}{6} = 3.5$$

It is important to not only know the mean or expected value of a stochastic variable but also the (expected) deviation from the mean value. As this deviation may be positive or negative, it is usually the square of the deviation that is taken, so $\{\underline{x} - E(\underline{x})\}^2$. The expectancy of this parameter is called the variance or the second moment:

$$\begin{aligned}\mathrm{var}(\bar{x}) &= E\left[\{\bar{x} - E(\bar{x})\}^2\right] = E\left[\bar{x}^2 - 2\bar{x}E(\bar{x}) + E^2(\bar{x})\right] \\ &= E\left(\bar{x}^2\right) - 2E(\bar{x})E(\bar{x}) + E^2(\bar{x}) \\ &= E\left(\bar{x}^2\right) + E^2(\bar{x})\end{aligned} \tag{2.23}$$

where E is supposed to be a linear operator. The square root of the variance is the standard deviation. This parameter has the same dimension as $\underline{x}$ itself, but it is not a stochastic variable.

Now we shall return to continuous variables. The first and second moments of a continuous stochastic variable are defined as:

$$E(x) = \int_{-\infty}^{\infty} xp(x)dx \tag{2.24}$$

$$\mathrm{var}(x) = \int_{-\infty}^{\infty} \{x - E(x)\}^2 p(x)dx \tag{2.25}$$

which are similar expressions to discrete variables. The expected value, in particular for the normal distribution function $x(t)$, is:

$$
\begin{aligned}
E(x) &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu) e^{-y^2/2\sigma^2} dy = \mu
\end{aligned}
\tag{2.26}
$$

The parameter μ in the Gauss distribution function is exactly the same as the expected value and it corresponds to the top of the probability density function (Figure 2.12). The variance is:

$$
\begin{aligned}
\operatorname{var}(x) &= E\left(x^2\right) - E(x)^2 \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(x-\mu)^2/2\sigma^2} dx - \mu^2 \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu)^2 e^{-y^2/2\sigma^2} dy - \mu^2 \\
&= \left(\sigma^2 + \mu^2\right) - \mu^2 = \sigma^2
\end{aligned}
\tag{2.27}
$$

The parameter σ in the normal distribution function appears to be precisely the standard deviation. It corresponds with the point of inflection in the probability density function (Figure 2.12).

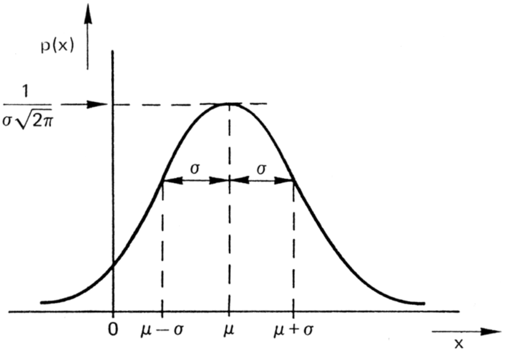


Figure 2.12. The probability density function of a normal or Gaussian distribution.

Next we shall relate the parameters μ and σ to some of the signal parameters introduced in 2.1.2. The mean or time average of a continuous signal was expressed as:

$$
x_m = \frac{1}{\tau} \int_0^{\tau} x(t) dt \tag{2.28}
$$

The time average of a stochastic continuous signal (such as a thermal noise voltage) is the same as its statistical mean value $E(x)$ provided that the statistical properties remain constant over the considered time interval. For signals with a normal distribution, this value equals μ, so $x_m = E(x) = \mu$. If, during the sampling or quantizing of the signal the statistic parameters do not change, the same will hold for discrete stochastic signals.
The power time average for a continuous signal $x(t)$ equals

$$P_m = \frac{1}{\tau}\int_0^{\tau} x^2(t)dt \tag{2.29}$$

(see Section 2.1.2). If the time average of the signal is zero, the mean power equals the square of the rms value. We then considered the time signal as a continuous stochastic variable. The variance appeared to be

$$\operatorname{var}(x) = \int_{-\infty}^{\infty} \{x - E(x)\}^2 p(x)dx \tag{2.30}$$

If $E(x) = 0$,

$$\operatorname{var}(x) = \int_{-\infty}^{\infty} x^2 p(x)dx \tag{2.31}$$

which is σ^2 for a normally distributed variable. For such signals, $P_m = \operatorname{var}(x) = \sigma^2$, if the statistical properties do not change during the time of observation. The rms value is therefore identical to the standard deviation σ (if the mean is zero).
Finally, there are some remarks to be made with respect to the relationship with the description in the frequency domain. The Fourier transform $F(\omega)$ of a time function $x(t)$ gives a complete description of a particular signal during its time of observation. Although it might well have identical statistical properties, another signal would result in another Fourier function. The conclusion is that the Fourier transform of $x(t)$ does not account for the statistical properties of the signal. The power density spectrum describes how the signal power is distributed over the frequency range. It can be proven that the power spectrum $S(f)$ is independent of a particular signal shape and is, therefore, a measure of the statistical properties of the signal.

SUMMARY

Signal description

- One possible categorizing of signals is:
  - static, quasi-static and dynamic,
  - deterministic and stochastic,

- continuous and discrete,
- analog and digital.

- Important signal characteristics are: peak value, peak-to-peak value, mean or time average value, root-mean-square (rms) value and mean power.
- Any periodic signal with period time T can be split up (expanded) into a series of sinusoids in which the frequencies are multiples of the fundamental frequency $f_0 = 1/T$ (Fourier series).
- The amplitudes of the sinusoidal components (Fourier coefficients) can be deduced from the signal time function using the Fourier transformation formulas (2.2). A periodic signal has a discrete spectrum or line spectrum while a non-periodic signal has a continuous spectrum.
- White noise is a noise signal that has a flat frequency spectrum over a wide frequency range.

Aperiodic signals

- The complex discrete Fourier series is given as $x(t)=\sum\limits_{n=-\infty}^{\infty} C_n e^{jn\omega t}$. A plot of C_n (the complex Fourier coefficients) versus frequency is the (complex) frequency spectrum.
- The Fourier transform $X(\omega)$ of $x(t)$ is: $X(\omega)=\int\limits_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$, the inverse transformation is $x(t)=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega$.
- The convolution of two signals $x_1(t)$ and $x_2(t)$ is defined as:
  $$g(\tau)=\int\limits_{-\infty}^{\infty} x_1(t)x_2(\tau-t)dt = x_1(t)*x_2(t)$$
- The Fourier transform of $x_1(t)*x_2(t)$ is $X_1(\omega)X_2(\omega)$, the Fourier transform of $x_1(t)x_2(t)$ is $X_1(\omega)*X_2(\omega)$.
- The frequency spectrum of a sampled signal with bandwidth B consists of multiple frequency bands positioned around multiples of the sampling frequency f_s. Each band is identical to the spectrum of $x(t)$, and is called an alias. If $f_s > 2/B$, the bands do not overlap.
- Shannon's sampling theorem, a highest frequency B signal, can be fully reconstructed after sampling if $f_s > 2/B$.
- The statistical properties of a stochastic signal are described by its distribution function $F(x) = P\{x(t) \le x\}$ and its probability density $p(x) = dF(x)/dx$.
- The expected value or first moment of a discrete stochastic variable is $E(\bar{x})=\frac{1}{N}\sum\limits_{i=1}^{N} p(x_i)x_i$, that of a continuous stochastic variable $E(x)=\int\limits_{-\infty}^{\infty} xp(x)dx$.
- The variance or second moment of a discrete stochastic variable is $\mathrm{var}(\bar{x})=E\left[\{\bar{x}-E(\bar{x})\}^2\right]=E\left(\bar{x}^2\right)+E^2(\bar{x})$; that of a continuous variable

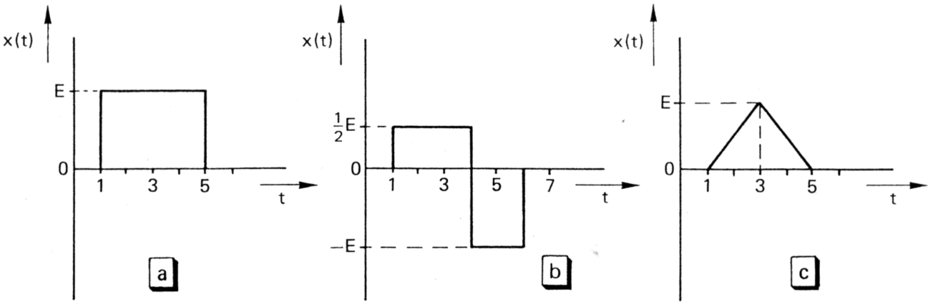
$\mathrm{var}(x)=\int_{-\infty}^{\infty}\{x-E(x)\}^2 p(x)dx$. The standard deviation is the square root of the variance.

- In the definition of the probability density of a normal or Gaussian distribution function $p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$, μ is the mean value and σ the standard deviation.
- The mean power of a signal with Gaussian amplitude distribution equals P_m = Var(x) = σ^2 while the rms value equals the standard deviation σ.

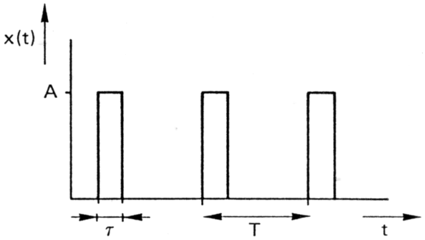
EXERCISES

Periodic signals

2.1 The figure below shows one period with three different periodic signals in period time T = 6 s. Find the peak-to-peak value, the time average and the rms value of all of these signals.



2.2 The crest factor of a signal is defined as the ratio between its peak value and its rms value. Calculate the crest factor of the signal below.



2.3 The signal given in Exercise 2.2 is connected to an rms voltmeter that can handle signals with a crest factor of up to 10. What is the minimum value of the signal parameter τ?

2.4 A voltmeter measures $|v|_m$ but is calibrated for the rms value of sinusoidal voltages. Predict the indication of this meter for each of the following signals:

a. a DC voltage of −1.5 V,

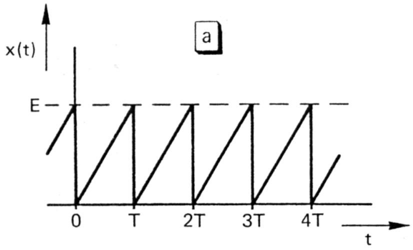
b. a sine voltage with an amplitude of 1.5 V,

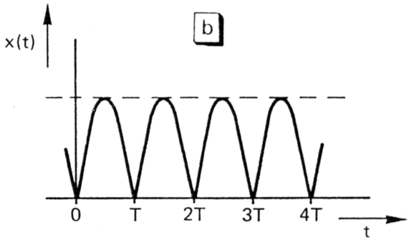
c. rectangular voltage (Figure 2.4a) with A = 1.5 V,

d. triangular voltage (Figure 2.4b) with $A = 1.5$ V.

2.5 An rms meter is connected to a signal source that produces a signal and noise. The indication appears to be 6.51 V. When the signal is turned off only the noise remains. The indication of the same meter then becomes 0.75 V. Calculate the rms value of the measurement signal without noise.

2.6 Find the Fourier coefficients of the following periodical signals, using Equations (2.3).

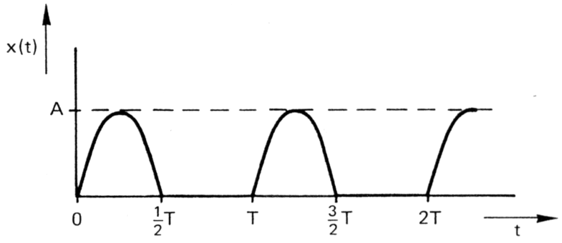




2.7 An electric resistor produces thermal noise with a spectral power density that is equal to $4kT$ (Johnson noise), k is Boltzmann's constant (1.38×10^{-23} J/K), T is the absolute temperature (K). Calculate the rms value of the noise voltage across the terminals of the resistor at room temperature (290 K) in a frequency range of 0 to 10 kHz.

Aperiodic signals

2.8 Find the frequency spectrum (amplitude and phase diagrams) of the signal given below (single-sided rectified sine)



2.9 A signal $x(t)$ is characterized as:
$x(t) = e^{-\alpha t}$ for $t > 0$
$x(t) = 0$ for $t \underline{>} 0$

a. Prove that the Fourier transform of $x(t)$ exists,
b. Determine the Fourier transform,
c. Draw the amplitude and phase spectrum.

2.10 Draw the distribution function $F(\underline{v})$ and the probability density function $p(\underline{v})$ of a stochastic signal with the following properties:
$\underline{v} = -5$ V, probability 0.2,
$\underline{v} = 0$ V, probability 0.5,
$\underline{v} = +5$ V, probability 0.3.

2.11 A signal $x(t)$ with Gaussian amplitude distribution has zero mean value. Calculate an expression for the expected value $E(y)$ of a signal $y(t)$ that satisfies:

$y(t) = x(t)$ for $x > 0$,
$y(t) = 0$ for $x \leq 0$
(single-sided rectified signal).

3 Networks

This chapter provides a brief introduction to the theory of networks. The first part will focus on electrical networks which are composed of electric network elements. The theory in question can also be applied to networks that consist of non-electrical components, as will be demonstrated in the second part of the chapter.

3.1 Electric networks

The main information carriers in an electronic measurement system are voltages and currents. These signals are processed by electronic components which are arranged and connected in such a way that the system is able to carry out the desired processing.

We may distinguish between passive and active components. It is the active components that make signal power amplification possible. The energy required is drawn from an auxiliary source such as a battery or mains supplies. It is impossible to gain power from components that are merely passive. Such components may store signal energy but they can never supply more energy than they have stored. Examples of passive components are resistors, capacitors, inductors and transformers. An example of an active component is the transistor. With semiconductor technology it is possible to integrate many transistors and other components so that the electronic building blocks (IC's or integrated circuits) become very compact. Such an IC may be seen as a single yet sometimes very complex electronic component, such as an operational amplifier or a microprocessor.

Electronic systems, circuits and components are modeled by networks consisting of the necessary network elements. Figure 3.1 summarizes all the existing electronic elements and indicates the corresponding relationships between the currents and voltages.

In the section above we have explicitly distinguished between components and elements. Network elements are models for particular properties of physical components. A capacitor, for instance, has capacitance C, but also dielectric losses modeled on the basis of a parallel resistance R. Similarly, an inductor not only has self-inductance L, but also resistance (i.e. wire resistance) and capacitance (between the wires). The properties of transistors can be described simply on the basis of current sources or voltage sources. A proper characterization would, however, require a more extended model including resistances and capacitances.

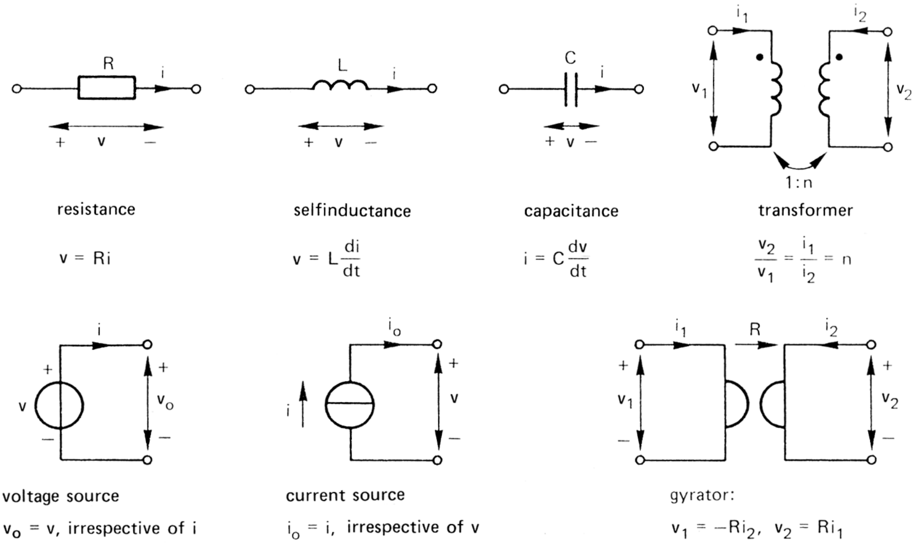

R
i
+ v −
resistance
$v = Ri$
L
i
+ v −
selfinductance
$v = L\frac{di}{dt}$
C
i
+ v −
capacitance
$i = C\frac{dv}{dt}$
i_1
i_2
v_1
v_2
1:n
transformer
$\frac{v_2}{v_1} = \frac{i_1}{i_2} = n$
i
+
v
−
v_o
voltage source
$v_o = v$, irrespective of i
i_o
i
+
v
−
current source
$i_o = i$, irrespective of v
i_1
R
i_2
v_1
v_2
gyrator:
$v_1 = -Ri_2$, $v_2 = Ri_1$

Figure 3.1. All electric network elements.

When modeling an electronic system composed of several components the corresponding network elements are connected to each other to form a complete network model of the system. A network contains nodes, branches and loops (Figure 3.2). A branch contains the network element, a node is the end of a branch and a loop is a closed path following arbitrary branches.

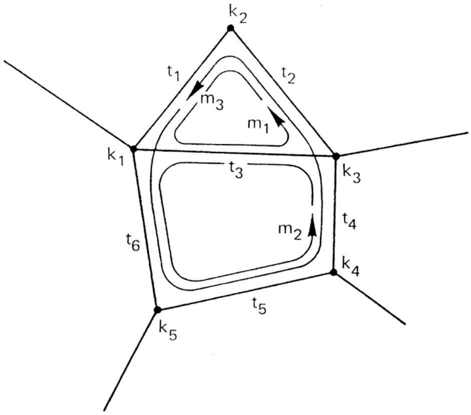

k_2
t_1
t_2
m_3
m_1
k_1
t_3
k_3
t_4
m_2
t_6
k_4
t_5
k_5

Figure 3.2. Nodes (k_i), loops (m_i) and branches (t_i) in a network.

Within networks it is Kirchhoff's rules that apply:

- the rule for currents: the sum of all currents flowing towards a node is zero: $\sum_k i_k = 0$

- the rule for voltages: the sum of all voltages along a loop is zero: $\sum_m v_m = 0$

All the voltages and currents in a network can be calculated using the voltage-current relations of the individual network elements and Kirchhoff's rules.

Example 3.1

The node voltage v_k in the network given in Figure 3.3 can be presented as a function of the voltages v_1, v_2 and v_3 at the end points of the elements R, L and C (all voltages relate to a common reference voltage). We can thus define the three currents: i_1, i_2 and i_3 (all of which are positive in the direction of the arrows). According to Kirchhoff's rule for currents $i_1 + i_2 + i_3 = 0$. Furthermore, we can apply the voltage-current relations of the three elements:

$$i_1 = \frac{1}{R}(v_1 - v_k)$$

$$i_2 = C\frac{d(v_2 - v_k)}{dt}$$

$$i_3 = \frac{1}{L}\int (v_3 - v_k)dt$$

If the three currents are eliminated from these four equations this will result in:

$$v_k + \frac{L}{R}\frac{dv_k}{dt} + LC\frac{d^2 v_k}{dt^2} = v_3 + \frac{L}{R}\frac{dv_1}{dt} + LC\frac{d^2 v_2}{dt^2}$$

The current direction can be randomly chosen. If, for instance, i_2 in Figure 3.3 were selected as positive in the opposite direction, the result would remain the same: the two equations with i_2 change into $i_1 - i_2 + i_3 = 0$ and $i_2 = C.d(v_k - v_2)/dt$.

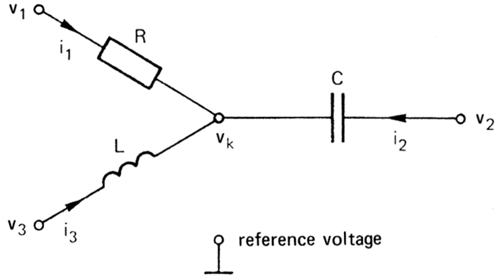


Figure 3.3. An example of the application of Kirchhoff's rule for currents.

In a network the relationships between voltages and currents are apparently expressed as differential equations. Differential equations have the following properties:

- they are linear: the highest power of the signal quantities and their time derivatives is 1. This is a result of the linear voltage-current network element relations. Linearity involves using superposition to facilitate calculations. In a network with several sources a current or voltage is found by separately calculating the contribution derived from each such source.
- they have constant coefficients which do not change over the course of time because it is assumed that the parameters of the network elements are constant

(like the resistance value). Linearity and coefficient constancy lead to the preservation of frequency: sinusoidal signals retain their shape and frequency.
- they are ordinary differential equations (not partial ones), time is the only independent variable.

These properties facilitate the solving of the differential equations by means of special calculation methods. Two of these methods are described in this book (see Chapter 4): they are based on complex variables and on the Laplace transform.

What determines the order of the system is the order of the differential equation. The differential equation given in the preceding example is a second order equation and so the network models a second order system.

The user of an electronic system is probably not interested in knowing all the system's voltages and currents. Only the signals on the terminals (the accessible points) will be of interest, for example the voltage between the output terminals of a transducer, or the input and output currents of a current amplifier. We think of the system as a closed box with a number of terminals. Its model (an electric network) should thus be conceived as a box as well: the only important nodes are those through which information exchange with the system's environment takes place. Depending on the number of external nodes, such a model will be called a two-terminal, three-terminal et cetera network (Figure 3.4). The terminals can be grouped in twos with each pair forming a port. A two-terminal network is therefore alternatively known as a one-port network and a four-terminal network as a two-port network, et cetera. Many electronic instruments and circuits have two ports with a common terminal that is usually grounded (zero potential) (Figure 3.5). The port to which the signal source is connected is the input port (or, in short: the input). The port from which the signal is taken is called the output port (or simply: the output). The corresponding voltages and currents constitute the input and output voltages (v_i and v_o) and the input and output currents (i_i and i_o) respectively. If we look at Figure 3.5 the following input and output relations can be distinguished: the voltage transfer (or voltage gain) v_o/v_i, current transfer $-i_o/i_i$, the voltage-to-current transfer i_o/v_i, the current-to-voltage transfer (or transconductance) v_o/i_i and the power transfer $p_o/p_i = -v_oi_o/v_ii_i$.

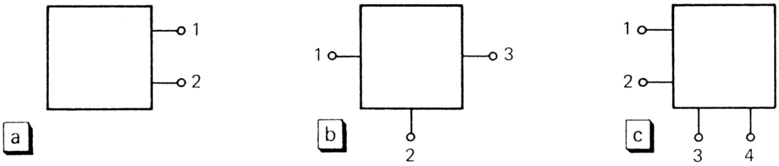


Figure 3.4. (a) a two-terminal network or one-port; (b) a three-terminal network; (c) a four-terminal network or two-port.

A network with n terminals can be fully characterized by a set of equations linking together all the external currents and voltages. Such an n-terminal network may be built up in many different ways but still characterized by the same set of equations. This property is used to break down electric networks into easily calculable structures.

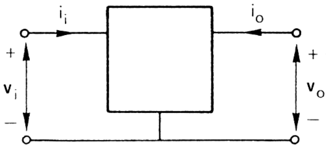


Figure 3.5. A three-terminal network connected as a two-port.

Example 3.2
Consider the network of parallel resistances in Figure 3.6a. For each branch k, $v = i_k R_k$. *According to Kirchoff's rule:* $i = i_1 + i_2 + i_3 + ... + i_n$. *Hence:*

$$\frac{i}{v} = \frac{1}{v}\sum_k i_k = \sum_k \frac{1}{R_k}$$

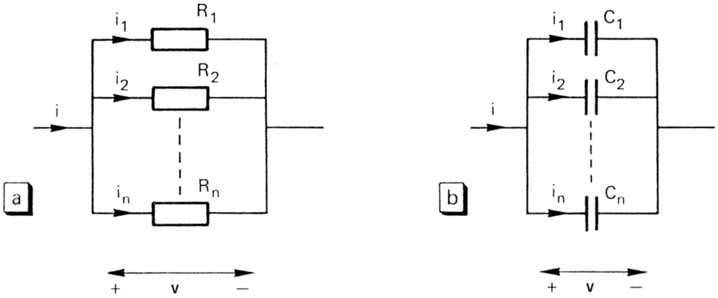


Figure 3.6. (a) network consisting of parallel resistances; (b) network composed of parallel capacitances.

A network consisting of parallel resistances is thus equivalent to single resistance in which the reciprocal value is $1/R_p = \sum_k 1/R_k$. *This value is always less than the smallest resistance.*

In the case of the network given in Figure 3.6b the equations $i = i_1 + i_2 + i_3 + ... + i_n$ *and* $i_k = C_k . dv/dt$ *apply, so*

$$i = \sum_k C_k \frac{dv}{dt}$$

The network with parallel capacitances is equivalent to a single value $C_p = \sum_k C_k$, *a value that is always larger than the largest capacitance.*

Example 3.3
Figure 3.7 displays two versions of a three-terminal network, both consisting of three resistances. The networks are equivalent for particular values of the resistances. To find the necessary conditions we must first calculate the resistance that can be measured between terminals 1 and 2 while leaving terminal three free (floating). When done for both networks the results must be equal, so:

$$R_1 + R_2 = \frac{R_{12}(R_{13} + R_{23})}{R_{12} + R_{13} + R_{23}}$$

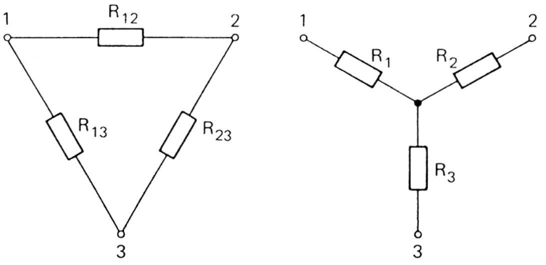


Figure 3.7. A triangular network can be converted into a star-shaped network and vice versa.

Similarly, two other relationships can be found for the resistances between terminals 2 and 3 and terminals 1 and 3. The conditions for equivalence can then be established on the basis of these three equations. The result will be:

$$R_i = \frac{R_{ij} R_{ik}}{R_{ij} + R_{ik} + R_{jk}} \qquad i, j, k = 1, 2, 3, \text{ cyclic.}$$

The formula for the inverse network transformation is:

$$R_{ij} = \frac{R_i R_j + R_i R_k + R_j R_k}{R_k} \qquad i, j, k = 1, 2, 3, \text{ cyclic.}$$

Example 3.4

Figure 3.8 shows a network with an input port and an output port.
To find the output voltage as a function of the input voltage, we assume that current i is flowing through the loop. The elimination of i from the equations $v_i = iR_1 + iR_2$ and $v_o = iR_2$ would result in:

$$v_o = v_i \frac{R_2}{R_1 + R_2}$$

It would appear that the output voltage is a fraction of the input voltage. This network is known as a voltage divider.

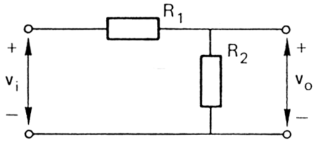


Figure 3.8. A voltage divider network.

3.2 Generalized network elements

In the first part of this chapter we revealed that the relations between electrical quantities in a network are expressed as differential equations. This is also true of many other physical systems. If we look at the system equations in various other disciplines, remarkable similarities can be seen. One might, for example, compare Ohm's law, $v = R \cdot i$, with the equation used to relate the heat current q through a thermal conductor to see the resulting temperature difference ΔT in that conductor:

$\Delta T = R_{th} \cdot q$, or one might take the equation that relates the force F on a mechanical damper to the speed v of the damper: $v = (1/b)F$. Similarly, analog forms of the equations $i = C(dv/dt)$ and $v = L(di/dt)$ can be found by describing the properties of mechanical, hydraulic, pneumatic and thermodynamic systems.

The quantities that describe technical systems appear to belong to one of two classes: through-variables and across variables. To explain this classification, we first need to introduce the term “lumped element”. A lumped element symbolizes a particular property of a physical component. It is imagined that in that element that property is concentrated between its two end points or nodes. Energy or information can only be exchanged through these terminals.

A through-variable is a physical quantity which is the same for both terminals in the lumped element. An across-variable describes the difference with respect to equal physical quantities at the terminals. In an electronic system, the current is a through-variable, while the voltage (or potential difference) is an across-variable. Just to simplify this concept we shall call a through-variable an I-variable and an across-variable a V-variable.

To indicate that a V-variable always relates to two points (nodes), we will give the subscripts a and b, hence V_{ab}.

A lumped element is described as a relation between one I-variable and one V-variable. There are three basic relations:

$$I = C\frac{dV_{ab}}{dt} \tag{3.1a}$$

$$V_{ab} = L\frac{dI}{dt} \tag{3.1b}$$

$$V_{ab} = RI \tag{3.1c}$$

In these equations, the parameters C, L and R stand for generalized capacitance, self-inductance and resistance. We will discuss each of these generalized network parameters separately.

- *Generalized capacitance*

The relationship between the I-variable and the V-variable for generalized capacitance is given in (3.1a). In the electrical domain, the relationship between the current and the voltage of a capacitance is described using the equation:

$$i = C\frac{dv_{ab}}{dt} \tag{3.2}$$

In the thermal domain the q heat flows towards a body and its temperature is given as:

$$q = C_{th}\frac{dT_{ab}}{dt} \tag{3.3}$$

Often, the reference temperature is 0 K, so $q = C_{th}(dT/dt)$. C_{th} is the heat capacitance of the body and it indicates the rate of temperature change at a particular heat flow speed. Newton's law of inertia, $F = m \cdot a$, can be rewritten as:

$$F = C_{mech} \frac{dv_{ab}}{dt} \tag{3.4}$$

Apparently, mass can be thought of as a “mechanical capacitance”.

- *Generalized self-inductance*

The ideal, generalized self-inductance is described in equation (3.1b). In the electric domain:

$$v_{ab} = L \frac{di}{dt} \tag{3.5}$$

A mechanical spring is described in an analog way:

$$v_{ab} = \frac{1}{k} \frac{dF}{dt} \tag{3.6}$$

with F being the force on the spring, v_{ab} the difference in speed between the two end points, and k the stiffness. Likewise, the equation for a torsion spring is:

$$\Omega_{ab} = \frac{1}{K} \frac{dT}{dt} \tag{3.7}$$

Here Ω_{ab} is the angular velocity, T the moment of torsion and K the stiffness of rotation.
The thermal domain lacks an element that behaves analogously to self-inductance.

- *Generalized resistance*

The relationship between the I and V-variables of a generalized resistance is given in 3.1c. In the electrical domain this is equivalent to Ohm's law:

$$v_{ab} = Ri \tag{3.8}$$

A mechanical damper is described analogously:

$$v_{ab} = \frac{1}{b} F \tag{3.9}$$

as mentioned before. The thermal resistance R_{th} is defined as the ratio between temperature difference and heat flow. The hydraulic resistance is described as the ratio between pressure difference and mass flow, et cetera.

There are even more similarities within the various groups of network elements. These are connected with the stored energy and the dissipated energy. In a generalized capacitance, the V-variable is responsible for the energy storage:

$$E = \int P dt = \int VI dt = \tfrac{1}{2} C V_{ab}^2 \tag{3.10}$$

Replacing C with, for instance, mass m of a moving body will result in the equation $E = \frac{1}{2} m v^2$, the kinetic energy. There is, however, one exception: thermal capacitance. In the thermal domain, the I-variable is a power quantity: q (W, J/s). The thermal energy is:

$$\int_{t_1}^{t_2} q dt = C_{th} T \tag{3.11}$$

In generalized self-inductance the I-variable accounts for the energy storage:

$$E = \int P dt = \int VI dt = \tfrac{1}{2} L I^2 \tag{3.12}$$

It then immediately follows that the energy stored in a torsion spring is: $\frac{1}{2}(1/K)T^2$. The energy stored in pure C and L-elements can be totally retrieved. This is not the case with R-elements which convert the electrical energy into thermal energy (heat). That is why they are called dissipating elements. The energy uptake amounts to:

$$P = V \cdot I = \frac{V^2}{R} = I^2 R$$

The energy uptake of other R-elements follows in the same way. Again, thermal resistance is an exception: the energy uptake is the I-variable q itself: $P = q = T/R_{th}$.
Any network that models a physical system in a particular domain can be transformed into a network for another domain, using the already presented analogies. The equations are the same so the calculation methods (to be discussed in ensuing chapters) will be applicable not only to electrical signals but also to other domains.

Example 3.5
The mercury reservoir of a thermometer has a (concentrated) heat capacitance C_k; the heat resistance of the glass wall is R_g. Furthermore, the temperature of the measurement object is T_a (relative to 0 K) and the temperature of the mercury is T_k. The (electric) model of this thermometer measuring the temperature of a gas, is depicted in Figure 3.9a.
The model can be extended to account for the heat capacity of the glass reservoir and the heat transfer coefficient between the glass and the measurement object (the surrounding gas): k (W/m^2K), see Figure 3.9b. The thermal resistance between the glass and the gas equals k·A, A being the contact area between the glass and the gas.

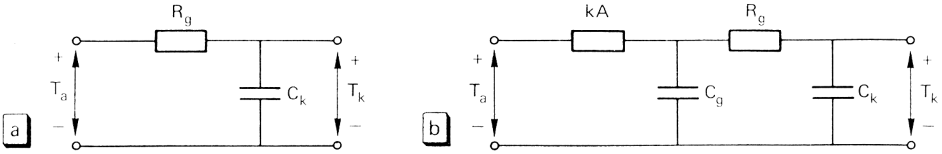


Figure 3.9. (a) A simplified electric analog model of a mercury-in-glass thermometer; (b) the model of the same thermometer, extended with the thermal capacity of the glass.

SUMMARY

Electric networks

- The most important network elements are: resistance, capacitance, self-inductance, current source and voltage source. The respective voltage-to-current relations for resistance, capacitance and self-inductance are $v = Ri$, $i = Cdv/dt$ and $v = Ldi/dt$.
- An electric network contains nodes, branches and loops. According to Kirchoff's rules, the sum of all currents leading to a node is zero, and the sum of all voltages around a loop is also zero.
- In an electric network currents and voltages are linked through ordinary, linear differential equations. The order of the system will correspond to the order of the differential equation.
- A number of resistances connected in series is equivalent to a single resistance, the value of which is the sum of the individual resistance values. This summing rule also applies to self-inductances in series and to capacitances in parallel, as well as to voltage sources in series and current sources in parallel.
- A number of resistances connected in parallel is equivalent to a single resistance, the reciprocal value of which is the sum of the reciprocal values of the individual resistances. This reciprocal summing rule also applies to self-inductances in parallel and to capacitances in series.

Generalized network elements

- Variables are divided into through-variables or I-variables and across-variables or V-variables.
- A lumped element is a model of a physical property which is thought to be concentrated between two terminals.
- A lumped element is characterized by a connection between an I-variable and a V-variable. There are three basic equations, corresponding to the three basic generalized elements of capacitance, self-inductance and resistance:

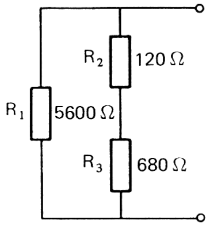
$$I = C\frac{dV}{dt};\ V = L\frac{dI}{dt};\ V = RI.$$

- The energy stored in generalized capacitance is $\frac{1}{2}CV^2$ and in generalized self-inductance it is $\frac{1}{2}LI^2$. The energy dissipated in a resistive or dissipative element is VI.
- There is no thermal self-inductance.

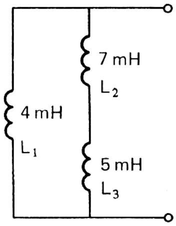
EXERCISES

Electric networks

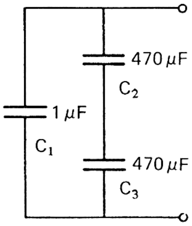
3.1 Replace each of the a-f two-terminal networks given below with a single element and calculate their equivalent values.


R₂
120 Ω
R₁
5600 Ω
R₃
680 Ω

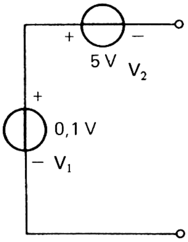
a


7 mH
L₂
4 mH
L₁
5 mH
L₃

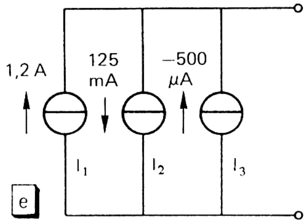
b


470 μF
C₂
1 μF
C₁
470 μF
C₃

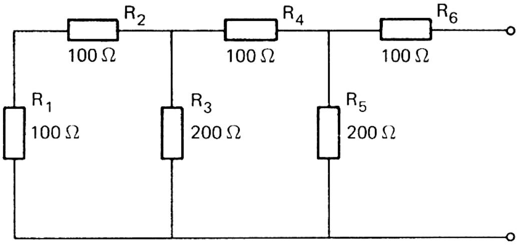
c



d


1,2 A
125 mA
−500 μA
I₁
I₂
I₃

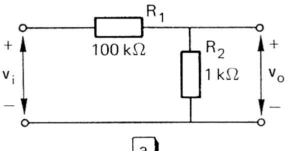
e


R₂
100 Ω
R₄
100 Ω
R₆
100 Ω
R₁
100 Ω
R₃
200 Ω
R₅
200 Ω

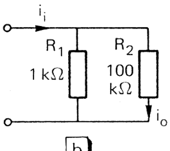
f

3.2 The resistance of one edge of a cube made up of 12 wires (forming the edges) is just 1 Ω. Calculate the resistance between the two end points of the cube's diagonal.

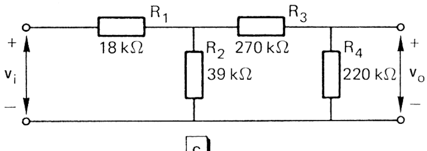
3.3 Calculate the a-c network transfer given below.


R₁
100 kΩ
R₂
1 kΩ
+
vᵢ
−
+
vₒ
−

a

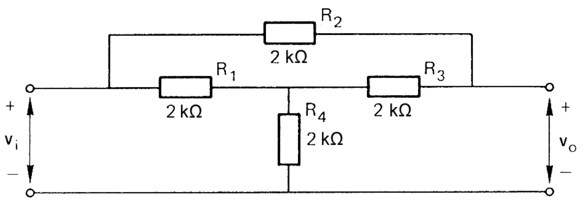

iᵢ
R₁
1 kΩ
R₂
100 kΩ
iₒ

b

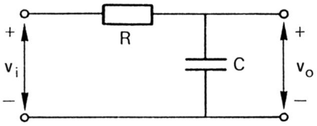

R₁
18 kΩ
R₃
270 kΩ
R₂
39 kΩ
R₄
220 kΩ
+
vᵢ
−
+
vₒ
−

c

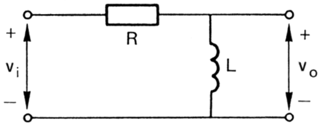
3.4 Calculate the network transfer shown below using the transformation formulas for triangular and star-shaped networks.



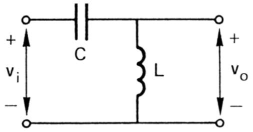
3.5 Work out the differential equations that describe the voltage transfer of the a-f networks depicted below.



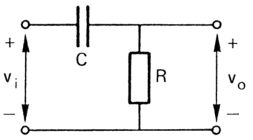
a



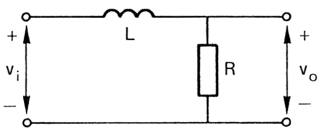
b



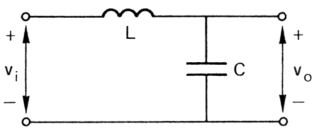
c



d



e



f

3.6 An ideal capacitor with capacitance C is charged from a constant current of 1 μA, starting from 0 volt. After 100 s the voltage is 20 V. Find the capacitance C.

Generalized network elements

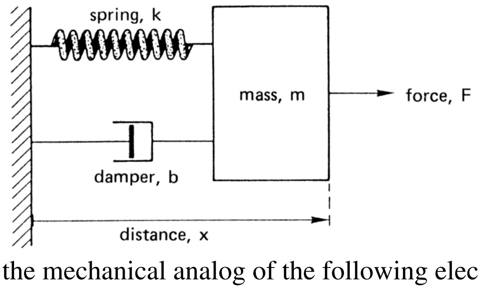
3.7 Which of the following elements or system properties can be described according to generalized capacitance, self-inductance and resistance:

thermal capacitance, mass, mechanical damping, moment of inertia, stiffness, thermal resistance?

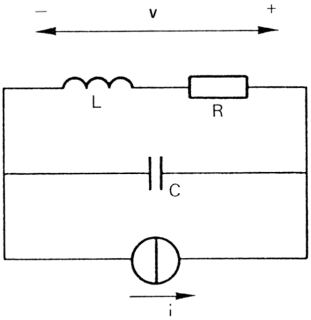
3.8 In addition to V and I-variables there are also time-integrated variables, for instance $x = \int v dt$. Assign one of them: V-variable, I-variable, integrated V-variable and integrated I-variable to each of the quantities below:

force, angular velocity, electric charge, heat flow, angular displacement, mass flow, heat, temperature.

3.9 Work out the electric analog model for the mass-spring system given in the figure below. What is the connection between F and x?



3.10 Deduce what is the mechanical analog of the following electrical network.


–
v
+
L
R
C
i

4 Mathematical tools

In order to compute the voltages and currents of an electric network one first has to solve a set of differential equations, an activity which – even with relatively simple networks – can be rather time-consuming. In this chapter two ways of facilitating the computations will be discussed. In the first part, we shall introduce the complex variables which can be used as a mathematical tool to calculate currents and voltages in a network without having to determine and solve the differential equations. The method is simple, but only valid for sinusoidal signals. In the second part, the Laplace transform, which can be used as a mathematical tool when computing arbitrary signals, will be introduced.

4.1 Complex variables

4.1.1 The properties of complex variables

This chapter commences with a brief overview of the main properties of complex variables. A complex variable is defined as the sum of a real variable and an imaginary variable. The latter is the product of a real variable and the imaginary unit $I = \sqrt{-1}$. To avoid confusion with the symbol i which stands for electric current, electrical engineers quickly adopted the symbol j instead: $j = \sqrt{-1}$. A complex variable is given as $z = a + jb$, in which a and b are real. The variables a and b are termed the real and imaginary components of the complex variable z. They are alternatively denoted as Re z and Im z, respectively. A complex variable can therefore be presented as $z = \mathrm{Re}\, z + j\, \mathrm{Im}\, z$.

Real variables are presented as points on a straight line. Complex variables are presented as points in a complex plane, their coordinates being positioned on the real and imaginary axes (Figure 4.1).

This same figure provides another representation of z, using a length and an angle as the two coordinates (polar coordinates). The distance between z and the origin (0,0) is the modulus or absolute value of z which is denoted as $|z|$. The angle between the "vector" and the positive real axis is the argument z, denoted as arg z, or simply given as a symbol for an angle, such as φ. The relationship between these two representations emanates directly from Figure 4.1:

$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2} \tag{4.1}$$

$$\varphi = \arg z = \arctan \frac{\operatorname{Im} z}{\operatorname{Re} z}$$

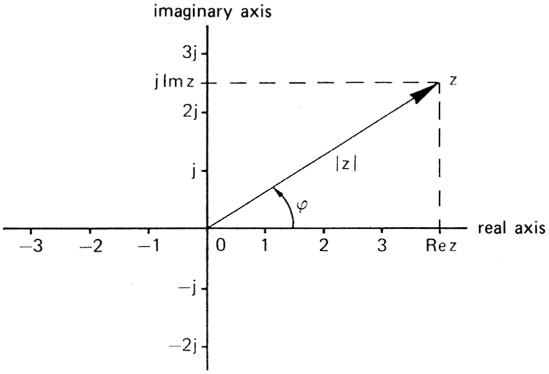


Figure 4.1. A complex variable is represented as a point in the complex plane.

Using the equations $\operatorname{Re} z = |z|\cos\varphi$and $\operatorname{Im} z = |z|\sin\varphi$, the complex variable can also be given as:

$$z = |z|(\cos\varphi + j\sin\varphi) \tag{4.2}$$

A third way to represent a complex variable is as

$$z = |z|e^{j\varphi} \tag{4.3}$$

From the definitions of complex variables the following rules for the product and the ratio of two complex variables can be derived:

$$\begin{aligned} &|z_1 z_2| = |z_1||z_2|; \quad \left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}; \\ &\arg z_1 z_2 = \arg z_1 + \arg z_2; \quad \arg\frac{z_1}{z_2} = \arg z_1 - \arg z_2 \end{aligned} \tag{4.4}$$

Furthermore, in the case of complex variables it is the common rules for integration and differentiation that hold. These rules will be frequently referred to in the coming chapters.

4.1.2 The complex notation of signals and transfer functions

What really characterizes a sinusoidal signal is its amplitude, frequency and phase. When such a sine wave signal passes through an electronic network its amplitude and phase may change but its frequency will remain unchanged.

There are marked similarities between the amplitude $\hat{x}$ and phase φ of sinusoidal signals $x(t) = \hat{x}\cos(\omega t + \varphi)$ and the modulus and argument of complex variables. The complex variable $X = |X|e^{j(\omega t + \varphi)}$ is represented in the complex plane as a rotating vector with length $|X|$ and angular speed ω (compare Figure 4.1). For $t = \pm\, nT$ the argument of X equals φ. Thus the modulus $|X|$ is equivalent to the amplitude $\hat{x}$ and the argument $\arg X$ is equivalent to the phase φ. In addition, the real part of X is just equal to the time function $x(t)$.

To distinguish between complex variables and real or time variables, the former are written in capitals (X, V, I), while the time variables are given in lower-case letters (x, v, i).

In Chapter 3 we defined several transfer functions as the ratio between output quantity and input quantity. Complex transfer functions can be similarly defined. For example, the complex voltage transfer function of a two-port network is denoted as $A_v = V_o/V_i$. The amplitude transfer follows directly from $|A_v|$ and the argument of A_v represents the phase difference between the input and output: $|A_v| = |V_o|/|V_i| = \hat{v}_o / \hat{v}_i$ and $\arg A_v = \arg V_o - \arg V_i = (\omega t + \varphi_o) - (\omega t + \varphi_i) = \varphi_o - \varphi_i$.

4.1.3 Impedances

The ratio of complex voltages and complex currents is generally a complex quantity. The ratio V/I is called impedance Z. The inverse ratio is the admittance: $Y = 1/Z = I/V$. Impedance can be viewed as complex resistance and admittance as complex conductance. We will now deduce the impedance of a capacitance and a self-inductance.

The voltage-current relationship for a self-inductance is $v = Ldi/dt$. A sinusoidal current can be represented as a complex current $I = |I|e^{j(\omega t + \varphi)}$. The complex voltage of the self-inductance is:

$V = LdI/dt = L|I|j\omega e^{j(\omega t + \varphi)} = j\omega LI$. This is the complex relation between the voltage and the self-inductance current. The impedance of the self-inductance becomes: $Z = V/I = j\omega L$.

The impedance of capacitance is found in a similar way. In the time domain, the current through the capacitance is $i = Cdv/dt$, so in complex notation: $I = CdV/dt = j\omega CV$. Hence, the impedance of capacitance is $Z = V/I = 1/j\omega C$.

From what has been stated above it follows that the impedances of self-inductance and capacitance have imaginary values. The impedance of a resistor is real. The impedance of a composition of network elements is, in general, a complex quantity. The same rules as those applied to series and parallel combinations, and to networks that only have resistances can be used to compute these impedances.

Example 4.1

The ratio of V and I in the network of Figure 4.2 is equal to:

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2} = \frac{\left(R_1 + 1/j\omega C\right)R_2}{R_1 + 1/j\omega C + R_2} = R_2 \frac{1 + j\omega R_1 C}{1 + j\omega\left(R_1 + R_2\right)C}$$

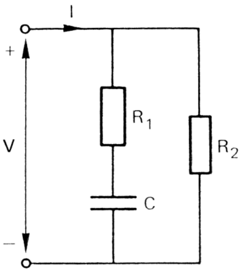


Figure 4.2. An example of a complex impedance.

The modulus of Z represents the ratio of the voltage amplitude and current amplitude: $|Z| = |V|/|I| = \hat{v}/\hat{\imath}$. The argument equals the phase difference between the sine-shaped voltage and the current.

The modulus of the impedance of self-inductance is $|Z| = \omega L$, which means that it is directly proportional to the frequency. The argument amounts to $\pi/2$. As $\arg V = \arg I + \arg Z = \arg I + \pi/2$, the current through self-inductance lags the voltage across it by $\pi/2$ radians.

The modulus of the impedance of a capacitance is $|Z| = 1/\omega C$ and so it is inversely proportional to the frequency. The argument is $-\pi/2$, so the current through the capacitance leads the voltage across it by $\pi/2$ radians. In composite networks, the phase difference is generally a function of frequency.

Example 4.2

The modulus $|Z|$ of the network given in Figure 4.2 is:

$$|Z| = \left|\frac{Z_1 Z_2}{Z_1 + Z_2}\right| = \frac{\left|R_1 + 1/j\omega C\right| R_2}{\left|R_1 + 1/j\omega C + R_2\right|} = R_2 \sqrt{\frac{1 + \omega^2 R_1^2 C^2}{1 + \omega^2\left(R_1 + R_2\right)^2 C^2}}$$

For $\omega \rightarrow 0$, $|Z|$ approaches R_2. This can also be instantly concluded from Figure 4.2, since at DC the capacitance behaves like an infinitely large resistance. In the case $\omega \rightarrow \infty$, the capacitance behaves like a short circuit for the signals, thus:

$$|Z(\omega \rightarrow \infty)| = \lim_{\omega \to \infty} R_2 \sqrt{\frac{1 + \omega^2 R_1^2 C^2}{1 + \omega^2\left(R_1 + R_2\right)^2 C^2}} = \lim_{\omega \to \infty} R_2 \sqrt{\frac{1/\omega^2 + R_1^2 C^2}{1/\omega^2 + \left(R_1 + R_2\right)^2 C^2}} = \frac{R_1 R_2}{R_1 + R_2}$$

which is nothing other than the two parallel resistances R_1 and R_2. The phase difference between the current through the network and the voltage across it emerges from:

$$\arg Z = \arctan \omega R_1 C - \arctan \omega\left(R_1 + R_2\right)C$$

Using complex expressions for the impedances of a self-inductance and a capacitance the transfer of two-port networks can be quite quickly achieved. The amplitude transfer and the phase difference can be immediately derived from the complex transfer.

Example 4.3
The complex transfer of the network given in Figure 4.3 is directly established from the formula for the voltage divider network
(see Exercise 3.3):

$$H = \frac{V_o}{V_i} = \frac{R_2}{R_2 + R_1 + 1/j\omega C} = \frac{j\omega R_2 C}{1 + j\omega(R_1 + R_2)C}$$

The modulus and the argument are:

$$|H| = \frac{\hat{v}_o}{\hat{v}_i} = \frac{\omega R_2 C}{\sqrt{1 + \omega^2 (R_1 + R_2)^2 C^2}}$$

$$\arg H = \frac{\pi}{2} - \arctan \omega(R_1 + R_2)C$$

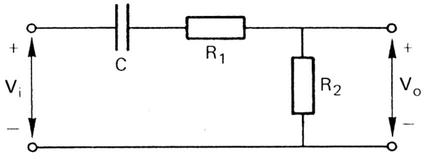


Figure 4.3. An example of a complex voltage transfer.

Both the amplitude transfer and the phase transfer are functions of frequency. In Chapter 6 we will introduce a simple method for plotting $|H|$ and arg H versus frequency in order to gain quick insight into the frequency dependence of the transfer.

4.2 Laplace variables

The Laplace transform is a well-known way of solving linear differential equations. Using the Laplace transform, a linear differential equation of order n is converted into an algebraic equation of order n. As electronic networks are characterized by linear differential equations, the Laplace transform may be a useful tool when it comes to describing the properties of such networks. The method is not restricted to sinusoidal signals, as is the case with complex variables. Laplace variables are valid for arbitrary signal shapes.

The Laplace transform can be seen as an extension of the Fourier transform (2.2.2). The conditions required for a Laplace transform are somewhat easier to fulfill (see Equation (2.7)). This aspect will be briefly discussed in Section 4.2.4.

4.2.1 The Laplace transform

The definition for the single-sided Laplace $x(t)$ function transform is:

$$X(p) = L\{x(t)\} = \int_0^\infty x(t)e^{-pt}dt \tag{4.5}$$

By means of this transformation a function $x(t)$ can be changed into a function $X(p)$ in the Laplace domain. The Laplace operator p is sometimes also represented by the letter s. In the double-sided Laplace transform the integration range goes from $-\infty$ to $+\infty$. Table 4.1 gives some of the time functions together with the corresponding Laplace functions. The time functions are presumed to be zero for $t < 0$.

Table 4.1. Some time functions with their corresponding Laplace transforms.

$x(t)$	$X(p)$	$x(t)$	$X(p)$
1	$\frac{1}{p}$	$t \cos \omega t$	$\frac{p^2 - \omega^2}{(p^2 + \omega^2)^2}$
$t^n \ (n \geq 0)$	$\frac{n!}{p^{n+1}}$	$t \sin \omega t$	$\frac{2p\omega}{(p^2 + \omega^2)^2}$
e^{at}	$\frac{1}{p-a}$	$e^{-at} \cos \omega t$	$\frac{p+a}{(p+a)^2 + \omega^2}$
$\cos \omega t$	$\frac{p}{p^2 + \omega^2}$	$e^{-at} \sin \omega t$	$\frac{\omega}{(p+a)^2 + \omega^2}$
$\sin \omega t$	$\frac{\omega}{p^2 + \omega^2}$	$\delta(t)$	1

When it comes to the transformation of other functions the following rules can be used. Let $L\{x(t)\} = X(p)$, then:

$$L\{ax(t)\} = aX(p) \tag{4.6}$$

$$L\{x_1(t) + x_2(t)\} = X_1(p) + X_2(p) \tag{4.7}$$

$$L\{e^{-at}x(t)\} = X(p+a) \tag{4.8}$$

$$L\{x(t-\tau)\} = e^{-p\tau}X(p) \tag{4.9}$$

$$L\left\{\frac{dx(t)}{dt}\right\} = pX(p) - x(0) \tag{4.10}$$

$$L\left\{\int x(t)dt\right\} = \frac{1}{p}X(p) \tag{4.11}$$

$$L\{x_1(t) * x_2(t)\} = X_1(p)X_2(p) \tag{4.12}$$

Successive repetition of the differentiation rule (4.5) results in:

$$L\left\{\frac{d^2x(t)}{dt^2}\right\} = p^2X(p) - px(0) - x'(0) \tag{4.13}$$

with $x'(0)$ the first derivative of $x(t)$ for $t = 0$.
In the next section we will see how the Laplace transform is used to solve network equations.

4.2.2 Solving differential equations with the Laplace transform

Chapter 3 showed that the relations between currents and voltages in an electric network are based on linear differential equations. To explain how the Laplace transform is used to solve such equations we shall now consider the network given in Figure 4.4a.

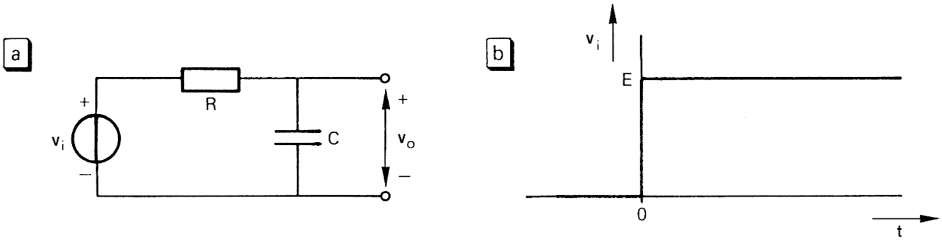


Figure 4.4. (a) An RC-network with (b) a step input voltage.

The relation between the output voltage v_o and the input voltage v_i is given as:

$$v_o + RC\frac{dv_o}{dt} = v_i \tag{4.14}$$

To solve this equation, we convert the time functions v(t) into Laplace functions $V(p)$. By using the rules given above and rearranging terms we can find that:

$$V_o(p) + pRCV_o(p) - RCv_o(0) = V_i(p) \tag{4.15}$$

This is a linear algebraic equation from which $V_o(p)$ can easily be solved:

$$V_o(p) = \frac{V_i(p) + RCv_o(0)}{1 + pRC} \tag{4.16}$$

If $v_i(t)$ is a known time function with an existing Laplace transform, $V_o(p)$ can be solved and by inverse transformation the output voltage $v_o(t)$ can finally be found. This procedure is illustrated using the input voltage shown in Figure 4.4b, a step function with height E. The output voltage is called the network step response.
The Laplace transform of the input appears to be E/p (see Table 4.1). Suppose all voltages are zero for $t < 0$, then:

$$V_o(p) = E\frac{1}{p(1+pRC)} \tag{4.17}$$

If we are to find the inverse transform of this function it must be split up into the terms listed in Table 4.1. This can be achieved by dividing the right-hand side of the last equation into terms with the respective denominators p and $1 + pRC$. This results in:

$$V_o(p) = E\left(\frac{1}{p} - \frac{1}{p + 1/RC}\right) \tag{4.18}$$

from which the output time function is found:

$$v_o(t) = E\left(1 - e^{-t/RC}\right) \quad t > 0 \tag{4.19}$$

Figure 4.5 shows this time function. It appears that the tangent at the point $t = 0$ intersects the horizontal line $v_o = E$ (the end value or steady-state value) for $t = \tau = RC$. With this property we can easily deduce the step response, provided that the value of RC is known. The product RC is called the time constant of the network, a term applicable to each first order system.

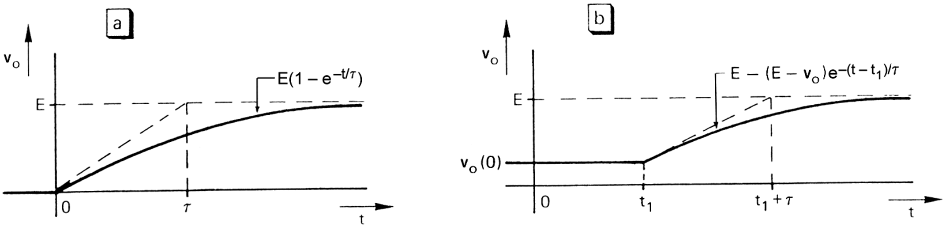


Figure 4.5. (a) Step response of the network from Figure 4.4a; (b) the step response of the same network, now with initial conditions and time delay.

If the capacitor was charged at $t = 0$, then $V_o(p)$ contains an additional term $v_o(0)/(p + 1/RC)$ which is why the expression for the output voltage given above must be extended with the term $v_o(t)e^{-t/RC}$.

If the input step starts at $t = t_1 > 0$ instead of at $t = 0$ the output voltage will be shifted over a time period of t_1 as well: in the expression for $v_o(t)$, t must be replaced by $t - t_1$. These two extra conditions, the initial charge and the time delay, are shown in Figure 4.5b. Notice that the tangent (now through the point $t = t_1$) still intersects the end value at time period τ after the input step.

4.2.3 Transfer functions and impedances in the p-domain

Generally the relationship between two signal quantities in a network is described using an n-th order differential equation:

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_0 y = b_m \frac{d^m x}{dt^m} + b_{m-1} \frac{d^{m-1} x}{dt^{m-1}} + \cdots + b_0 x \qquad (4.20)$$

If one applies the Laplace transform and supposes that the initial conditions are zero this changes into:

$$a_n p^n Y + a_{n-1} p^{n-1} Y + \cdots + a_0 Y = b_m p^m X + b_{m-1} p^{m-1} X + \cdots + b_o X \qquad (4.21)$$

The ratio between X and Y then becomes:

$$\frac{Y}{X} = \frac{b_m p^m + b_{m-1} p^{m-1} + \cdots + b_o}{a_n p^n + a_{n-1} p^{n-1} + \cdots + a_0} \qquad (4.22)$$

Equation (4.22) describes either a transfer function $H(p)$ or an impedance $Z(p)$, depending on the dimensions of X and Y. The Fourier transform $H(\omega)$ in Section 4.1 provides a description of the system in the frequency domain (for sinusoidal signals). Likewise, the transfer function $H(p)$ describes the properties of the system in the p-domain for arbitrary signals. This parallel with the Fourier transform also holds for impedances. In the p-domain, the impedance of a capacitance is $1/pC$ while that of a self-inductance is pL. The Fourier transform may be seen as a special case of the Laplace transform, namely $p = j\omega$, which means that only sinusoidal functions are considered.

Example 4.4

The transfer of the network depicted in Figure 4.4a can be written directly (without working out the differential equation) as:

$$H(p) = \frac{V_o(p)}{V_i(p)} = \frac{1/pC}{R + 1/pC} = \frac{1}{1 + pRC}$$

The impedance of a resistor, a capacitance and a self-inductance, all in series, amounts – in the p*-domain – to $R + 1/pC + pL$. The impedance of a network composed of a self-inductance that is in parallel to a capacitance is* $\frac{(1/pC)pL}{1/pC + pL} = \frac{pL}{1 + p^2 LC}$.

In Equation (4.22) the values of p for which the numerator is zero are called the zeroes of the system. The values of p for which the denominator is zero are termed the poles of the system. The transfer function of the network in Figure 4.4 has only one pole, $p = -1/RC$. The impedance of the network which consists of a parallel capacitance and self-inductance (see Example 4.4) has one zero for $p = 0$ and two imaginary poles $p = \pm j/\sqrt{LC}$.

The dynamic behavior of the system is fully characterized by its poles and zeroes. In some technical disciplines, in particular control theory, the description of systems is based on poles and zeroes.

4.2.4 The relation to the Fourier integral

The Fourier series involves expanding a periodic signal into discrete, sinusoidal components. The resultant Fourier integral may be viewed as an expansion into a continuous package of sinusoidal components. The Laplace operator p is a complex variable, $p = \alpha + j\omega$. The transform can therefore also be written as:

$$L\{x(t)\} = \int_0^\infty x(t)e^{-\alpha t}e^{-j\omega t}dt$$

This shows that the Laplace transform is equivalent to the Fourier transform (2.6), except that it is not the function $x(t)$ but rather the function $x(t)e^{-\alpha t}$ that is transformed. A proper choice of α allows functions that do not converge for $t \to \infty$ to be transformed, assuming that the function $x(t)e^{-\alpha t}$ satisfies the condition (2.7). In this respect, the Laplace transform can be interpreted as an expansion into a continuous package of signals of the type $e^{-\alpha t}\sin\omega t$, which are, in fact, exponentially decaying sine waves.

SUMMARY

Complex variables

- A complex variable z can be written as $z = \mathrm{Re}\, z + j\, \mathrm{Im}\, z = |z|(\cos\varphi + j\sin\varphi) = |z|e^{j\varphi}$, with $|z|$ the modulus and φ the argument of z.
- The modulus of z is $|z| = \sqrt{(\mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2}$, the argument of z is $\arg z = \arctan \mathrm{Im}\, z/\mathrm{Re}\, z$.
- The complex notation for a sinusoidal voltage or current $\hat{x}\cos(\omega t + \varphi)$ is $X = |X|e^{j(\omega t + \varphi)}$.
- The impedance Z of a two-terminal element is defined as the ratio between the complex voltage V and the complex current I. The impedance of a capacitance is $1/j\omega C$, that of a self-inductance $j\omega L$.
- The complex transfer H of a two-port network represents the ratio between the complex output signal and the complex input signal. The modulus $|H|$ represents the amplitude transfer while the argument $\arg H$ represents the phase transfer.
- Complex notation is only valid for sinusoidal signals.

Laplace variables

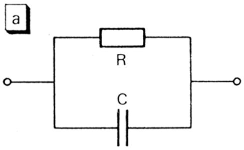
- The Laplace transform is defined as $X(p) = L\{x(t)\} = \int_0^\infty x(t)e^{-pt}dt$.
- The Laplace transform of the derivative of a function $x(t)$ is $L\{dx(t)/dt\} = pX(p) - x(0)$. From this property it follows that a linear differential equation can be converted into a linear algebraic equation using the Laplace operator p as the variable.

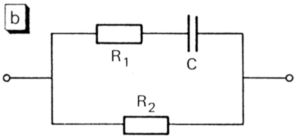
- The Laplace transform makes it possible to compute the properties of an electric network for arbitrary signals. The relevant complex notation is only applicable to sine waves.
- Transfer functions and impedances can be described within the p or Laplace domain. The impedance of a capacitance C and a self-inductance L is $1/pC$ and pL, respectively. The rules for the composition of networks in the ω-domain also apply to the p-domain.
- The zeroes of a system described using a Laplace polynome $T(p)/N(p)$ are those values of p for which $T(p) = 0$; the poles of this system are those values of p for which $N(p) = 0$.
- The Fourier integral can be described as the expansion of a function into a continuous series of sinusoidal components. Likewise, the Laplace integral can be seen as the expansion of a function into a continuous series of exponentially decaying sinusoids.

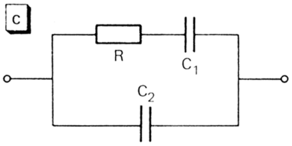
EXERCISES

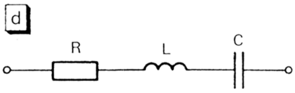
Complex variables

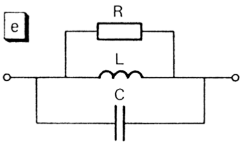
4.1 Find the impedance of each of networks a-f given below.

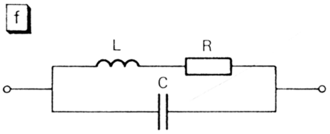






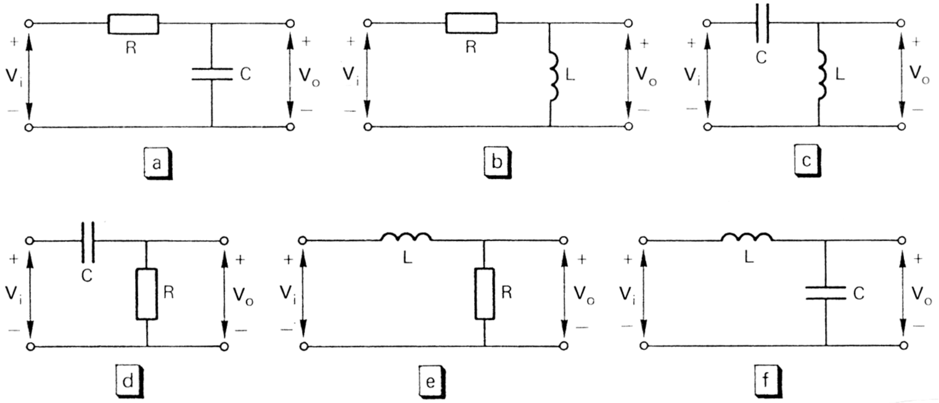




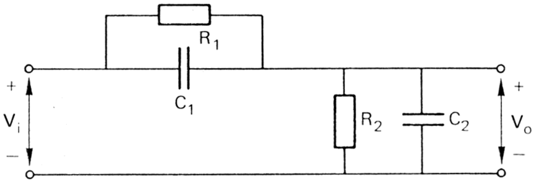


Calculate the impedance of network d for $\omega = 1/\sqrt{LC}$ and $R = 0$.
Calculate the impedance of network e for $\omega = 1/\sqrt{LC}$ and $R = \infty$.

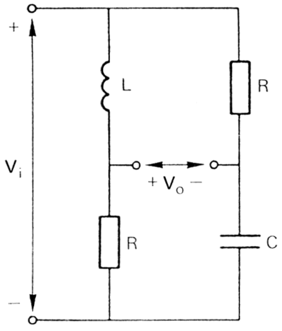
4.2 Find the complex transfer functions of networks a-f depicted below.



4.3 Find the complex transfer function of the network given below. Under which condition with respect to R_1, R_2, C_1 and C_2 is the transfer independent of the frequency?



4.4 Find the transfer function of the bridge network given below. Under which condition is $V_o = 0$, irrespective of V_i?



Laplace variables

4.5 Give the impedances in the p-domain of all the networks present in Exercise 4.1. All currents and voltages are zero for $t < 0$.

4.6 Transform the complex transfer functions of Exercise 4.2 into the p-domain. Find the zeroes and poles of these functions.

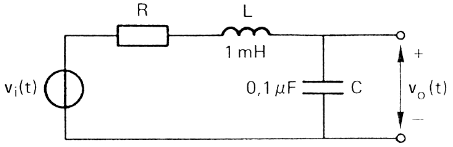
4.7 Using the Laplace transform calculate the output signal of the network given in Exercise 4.2d for each of the following situations:

a. $v_i = 0$ for $t < 0$, $v_i = E$ for $t \geq 0$, C uncharged at $t = 0$,

b. $v_i = 0$ for $t < 0$, $v_i = E$ for $0 \leq t < t_1$, $v_i = 0$ for $t > t_1$, C uncharged at $t = 0$.

Make a plot of the output voltage versus time for both situations.

4.8 In the network shown below, the input voltage $v_i(t)$ is a sine wave. At $t = 0$ the input is connected to the ground. At that moment, the voltage across capacitance C is just zero while the current through it is i_o.



Calculate the output voltage $v_o(t)$ for three values of R:

a. $R = 400\ \Omega$;

b. $R = 120\ \Omega$;

c. $R = 200\ \Omega$.

5 Models

Any model of an electronic measuring system should include properties that are of specific interest to the user. A more elaborate model is unnecessary and might even prove confusing, while a more restricted model would not provide sufficient information on system behavior and might even precipitate an incorrect interpretation of the measurement results.

The electronic properties of any measurement system can be modeled using a limited number of network elements (sources, impedances). The first part of this chapter will deal with the ways of obtaining such models and how to use them. The second part of the chapter will go on to illustrate how noise and interference signals are modeled.

5.1 System models

An electronic circuit is modeled according to a network of electronic elements. The model of a system with n external connections has (at least) n external terminals. In this chapter we will consider networks with only two, three or four terminals that are arranged either in one or two signal ports. What we shall particularly look at is the influence that this has on signal transfer when one system is being connected to another.

5.1.1 Two-terminal networks

In Chapters 3 and 4 we discussed two-terminal systems with passive elements. Such networks are equivalent if their impedances (as measured between the two terminals) are the same.

Such equivalence also exists for systems containing active elements, like voltage sources and current sources. Two (or more) active two-terminal networks are equivalent if the short-circuit current I_k and the open voltage V_o are equal (Figure 5.1). The ratio between the open voltage and the short-circuit current is the internal impedance or source impedance of the network: $Z_g = V_o/I_k$.

According to the theorem of Thévenin any active, linear two-terminal system can be fully characterized as having one voltage source V_o and one impedance, Z_g, connected in series (Figure 5.2a). An equivalent model consists of one current source I_k and one impedance Z_g (Figure 5.2b). This is known as Norton'sheorem.

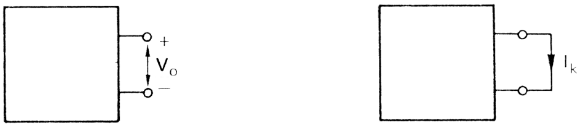


Figure 5.1. The determination of the open voltage and the short-circuit current of an active two-terminal network.

The equivalence between both models can easily be verified. With open terminals, the current through Z_g as depicted in Figure 5.2a is zero which means that the voltage across the terminals is just V_o. In Figure 5.2b all the current flows through Z_g when the terminals are open, so the output voltage equals $I_k \cdot Z_g = V_o$, which is the Thévenin voltage. Similarly, it can be shown that in both cases a current, I_k, flows through an external short-circuit.

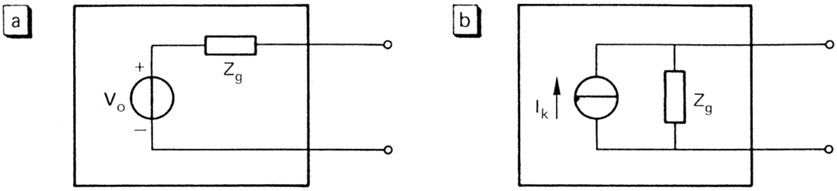


Figure 5.2. (a) Thévenin's equivalent of a two-terminal system; (b) Norton's equivalent.

Both models are completely equivalent but one may be preferred to the other depending on the properties of the system that needs to be modeled. Systems with a low source impedance value are preferably modeled according to the voltage source model while systems that behave more like a current source (with a high value of Z_g) are better if modeled according to the current source model. In such cases, the source impedance characterizes the deviation from an ideal voltage source ($Z_g = 0$) or, as the case might be, the current source ($Z_g = \infty$).

There are other possible reasons for making particular choices. For instance, a transducer with an output voltage that is proportional to the measurement quantity should preferably be modeled using a voltage source. A transducer that reflects the measurement quantity as a current is modeled using a current source model.

5.1.2 Two-port networks

Networks with three or four terminals generally have one port that is assigned to being the input port while the other becomes the output port (see also 3.1). In accordance with these functions we may define the input quantities V_i and I_i and the output quantities V_o and I_o. The polarity of these quantities is as is indicated in Figure 5.3.

The ratio between input voltage and input current is called the system's input impedance $Z_i = V_i/I_i$. The output impedance is the ratio between the output voltage and the output current: $Z_o = V_o/I_o$.

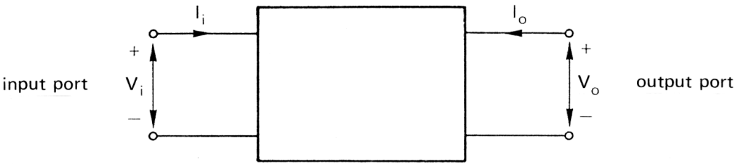


Figure 5.3. A two-port network with in- and output quantities.

There are many ways of modeling a two-port network. One obvious way is by expressing the output quantities in terms of the input quantities because generally the output is affected by the input. If the network contains no independent internal sources the system can be described in terms of the following two equations:

$$\begin{aligned} V_o &= AV_i + BI_i \\ I_o &= CV_i + DI_i \end{aligned} \tag{5.1}$$

where A, B, C and D are system parameters.
Although this is a rather logical way of describing a system there are other ways, for instance expressing both voltages in terms of currents, or both currents in terms of voltages.

$$\begin{aligned} V_i &= A'I_i + B'I_o \\ V_o &= C'I_i + D'I_o \end{aligned} \tag{5.2}$$

or

$$\begin{aligned} I_i &= A''V_i + B''V_o \\ I_o &= C''V_i + D''V_o \end{aligned} \tag{5.3}$$

The advantage of these latter relationships is that they can be directly transferred to an equivalent model (Figure 5.4). The system shown in Figure 5.3 is described according to two voltage sources, one at the input point and one at the output juncture (Figure 5.4a). In Figure 5.4a the system equations are:

$$\begin{aligned} V_i &= Z_{11}I_i + Z_{12}I_o \\ V_o &= Z_{21}I_i + Z_{22}I_o \end{aligned} \tag{5.4}$$

The coefficients Z_{11}, Z_{12}, Z_{21} and Z_{22} have a clear physical meaning (impedances) and can be directly determined on the basis of measurements. The two voltage sources in the network model are dependent sources; their value depends on other variables in the system.

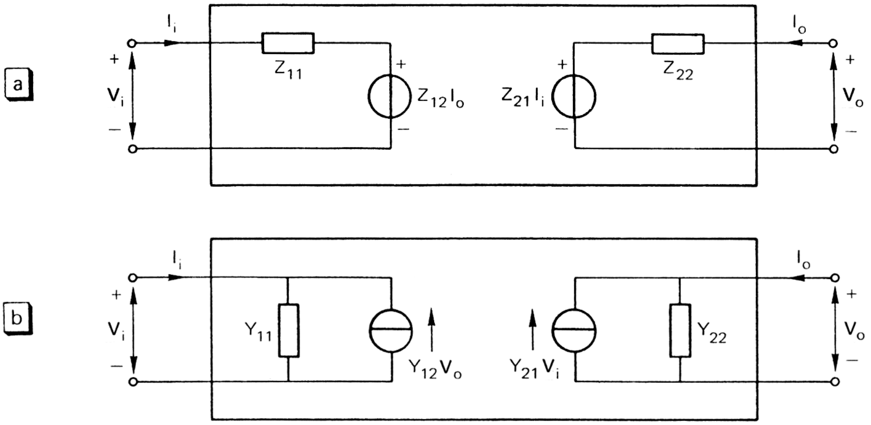


Figure 5.4. Various models for a two-port system: (a) with two voltage sources; (b) with two current sources.

In Figure 5.4b, both ports are modeled on a current source. The system equations are:

$$\begin{aligned} I_i &= Y_{11}V_i - Y_{12}V_o \\ I_o &= -Y_{21}V_i + Y_{22}V_o \end{aligned} \tag{5.5}$$

The Y system parameters can be converted into the Z parameters of the model in Figure 5.4a. Another possible model option is to have one voltage source and one current source.

It should be noted that, in general, the input impedance defined above is not equal to impedance Z_{11}, nor is output impedance Z_o equal to impedance Z_{22} in the model. Both Z_i and Z_o depend on the impedances that may be connected to the output or input port. It follows from the system equations or from the system models that input impedance Z_i in Figure 5.4a is only equal to Z_{11} if the output current $I_o = 0$ (open output) and that $Z_o = Z_{22}$ only if the input current $I_i = 0$ (open input).

Example 5.1

The system with the model depicted in Figure 5.5 is loaded with impedance Z_L at the output. Its input impedance, Z_i, is defined as V_i/I_i:

$$Z_i = \frac{V_i}{I_i} = \frac{I_iZ_{11} + I_oZ_{12}}{I_i}$$

with

$$I_o = \frac{-Z_{12}I_i}{Z_{22} + Z_L}$$

From these equations it follows that:

$$Z_i = Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22} + Z_L}$$

Similarly, when the output impedance of the same system has a source impedance connected at the input it is found to be:

$$Z_o = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11} + Z_g}$$

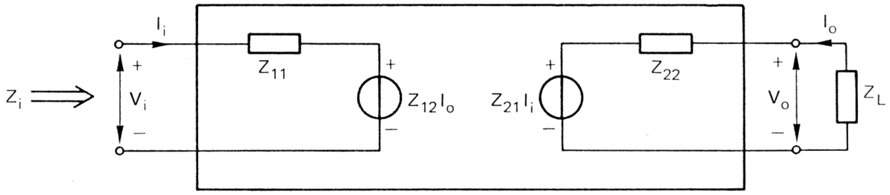


Figure 5.5. The input impedance of a system depends on the load at its output.

From Figure 5.4a it emerges that the output voltage which is $V_o = Z_{21}I_i + Z_{22}I_o$ is therefore a function of both the input current and the output current. Something similar goes for the input voltage which is: $V_i = Z_{11}I_i + Z_{12}I_o$. Apparently there is bi-directional signal transfer from input to output and vice versa. In many systems, like with amplifiers and measurement systems, only unidirectional signal transfer (from input to output) is allowed. Most measurement systems are designed to make reverse transfer negligible. In the model the source $Z_{12}I_o$ is therefore zero. One consequence of having a unidirectional signal path is that the input impedance becomes independent of what is connected to the output (the load) while the output impedance is independent of the source circuit connected to the input. Let us suppose, in Figure 5.5, that the impedance $Z_{12} = 0$, then $Z_i = Z_{11}$ and $Z_o = Z_{22}$. Such a system is fully described by its input impedance, its output impedance and a third system parameter (which is Z_{21} in the case of Figure 5.5).

Example 5.2

Figure 5.6 depicts the model of a voltage-to-current converter. The input and output have a common ground terminal, the input impedance is Z_i and the output impedance is Z_o. The reason for choosing the current output model is because the output should have the character of a current source. The value of the output current is S times the input voltage. S is the voltage-to-current sensitivity of the system.

There is no internal source on the input side of the model. It would seem that the input voltage and the current are not influenced by the output signals. In other words, there is no internal feedback.

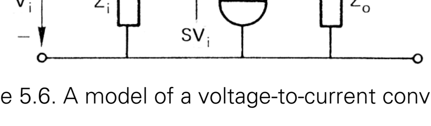


Figure 5.6. A model of a voltage-to-current converter.

5.1.3 Matching

In electronic systems the main carriers of information are currents and voltages. The proper transfer of information requires proper voltage or current transfer.

If voltage is the information carrier then the voltage transfer from one system to the other should be as accurate as possible in order to avoid information loss. If a signal source with internal source impedance Z_g is connected to a system whose input impedance is Z_i (Figure 5.7) the actual input voltage will be

$$V_i = \frac{Z_i}{Z_g + Z_i} V_g \tag{5.6}$$

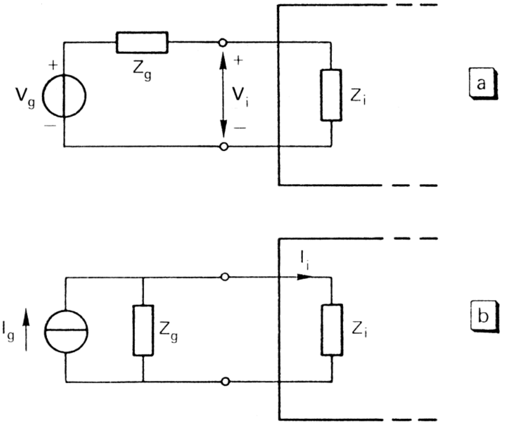


Figure 5.7. Matching between two systems; (a) voltage matching; (b) current matching.

This means that the input voltage is always smaller than the source voltage. The deviation is minimized by minimizing the ratio Z_g/Z_i. For $Z_g << Z_i$ the input voltage of the system is

$$V_i = \frac{1}{1 + Z_g/Z_i} V_g \approx \left(1 - Z_g/Z_i\right) V_g \tag{5.7}$$

so the relative error is about $-Z_g/Z_i$. Only if Z_g and Z_i are both known can the measurement error for this additional signal attenuation be corrected. Otherwise the input impedance of a voltage-measuring system should be as high as possible in order to minimize the load error. This process is known as the voltage matching of the systems.

If the current is the information carrier then the model given in Figure 5.7b will be preferred. The current I_i through the system's input is:

$$I_i = \frac{Z_g}{Z_g + Z_i} I_g \tag{5.8}$$

Here, too, the signal is attenuated. If $Z_i << Z_g$, the input current will be

$$I_i = \frac{1}{1 + Z_i/Z_g} I_g \approx \left(1 - Z_i/Z_g\right) I_g \tag{5.9}$$

and the relative error will be about $-Z_i/Z_g$. Again, to minimize this error, the measurement system's input impedance must be as low as possible. This is termed system current matching.

The results summarized:

- in order to minimize loading errors, voltage measurement systems require a high input impedance;
- what a current measurement system requires is a low input impedance.

When voltage or current matching is perfect, the signal transfer from the source to the input terminals of the measurement system is just 1; the power transfer, however, is zero. This is because when $Z_i = 0$, the input voltage is zero and with $Z_i \to \infty$ the input current is zero. In both cases, the input power is therefore also zero. Since a tremendous amount of signal power is required to activate most output transducers this might, at first sight, seem to be an undesirable situation. Fortunately, though, electronic components offer an almost infinite supply of signal power amplification, so voltage or current matching does not necessarily lead to low power transfer.

Another thing that current or voltage matching is responsible for is the zero power supplied by the signal source. This can be a great advantage, especially when the input transducer is unable to supply enough power by itself, or when no power may be extracted from the measurement object (for instance so as not to reduce measurement accuracy).

Voltage or current matching is not always possible. At high frequencies it is particularly difficult to realize high input impedance. Any system has an input capacitance that is different from zero and originates from the input components as well as from the connector and the connecting wires. The impedance of a capacitance decreases as frequency increases and so does the system's input impedance.

Another point to bear in mind is that high frequency signals may reflect at the interfaces between system parts. Such reflecting may introduce standing waves between two points on the signal path that will then interfere with the proper propagation of the measurement signal.

Such effects can be avoided by introducing another type of matching known as characteristic matching (see Figure 5.8). The systems in question have a particular input and output impedance which is the same for all the systems involved. This so-called characteristic impedance R_k has a fixed, relatively low and real value of, for instance, 50 or 75 Ω.

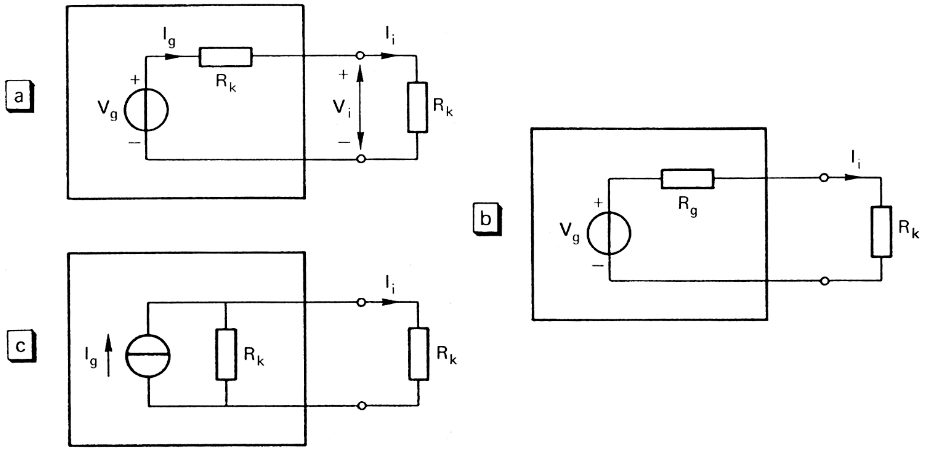


Figure 5.8. (a) characteristic matching of the signal source;
(b) the source supplies the maximum signal power to the system if $R_g = R_k$;
(c) characteristic loading of a current source.

The voltage, current and power signal transfer of a typical system differ essentially from those of the systems described so far. The voltage transfer V_i/V_g of the system shown in Figure 5.8a is just $\frac{1}{2}$. To calculate the power transfer from one system to another the reader is referred to Figure 5.8b which shows a voltage source with source impedance R_g that is loaded with impedance R_i. The signal power supplied to the load is:

$$P_i = I_i^2 R_i = \frac{V_g^2 R_i}{\left(R_g + R_i\right)^2} \tag{5.10}$$

and depends on both R_i and R_g. To find the conditions for maximal power transfer we must take the first derivative of P_i to the variable R_i:

$$\frac{dP_i}{dR_i} = V_g^2 \frac{\left(R_g + R_i\right)^2 - 2R_i\left(R_g + R_i\right)}{\left(R_g + R_i\right)^4} \tag{5.11}$$

This is zero for $R_i = R_g$, so maximum power transfer occurs when the two systems are characteristically coupled. Half of the available power is dissipated in the source resistance: $I_g^2 R_k = V_g^2/4R_k$. The other half is transferred to the load: $V_i{\cdot}I_i = V_g^2/4R_k$. The same conclusions can be drawn for the current source given in Figure 5.8c. This type of matching is also called characteristic matching or power matching because the power transfer is maximized.

Many characteristic systems only have a characteristic input impedance if the system is loaded with R_k, and characteristic output impedance if the system is connected to a source where the source resistance is just R_k (Figure 5.9). Such systems display bi-directional signal transport.

When using calibrated characteristic systems, care should be taken to connect characteristic impedances on both the input and output sides, otherwise transfer will deviate from the specified (calibrated) value.

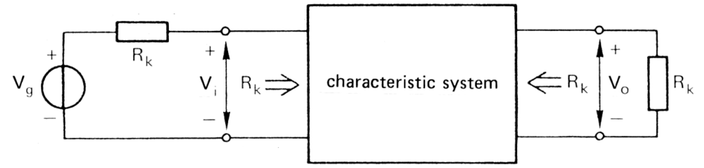


Figure 5.9. Some systems have only a characteristic input impedance if loaded with a characteristic impedance, and vice versa.

5.1.4 Decibel notation

Let us now consider two systems, connected in series. Let system 1 be characterized as $x_{o1} = H_1 x_{i1}$ and system 2 as $x_{o2} = H_2 x_{i2}$. These systems are coupled in such a way that $x_{o1} = x_{i2}$. The output of the total system equals $x_{o2} = H_1 H_2 x_{i1}$. Obviously, the total transfer of a number of systems connected in series is equal to the product of the individual transfers (taking into account the corrections made for non-ideal matching).

In several technical disciplines (telecommunications, acoustics) it is common to express the transfer as the logarithm of the ratio between output and input quantity. This simplifies the way in which the transfer of cascaded systems is calculated. The logarithmic transfer of the total system is simply the sum of the individual transfers based on the rule $\log(a \cdot b) = \log(a) + \log(b)$.

The logarithmic power transfer is defined as $\log(P_o/P_i)$, base 10, (unit bel), or as $10 \cdot \log(P_o/P_i)$ (unit decibel or dB).

In Figure 5.10 the system's input power is $P_i = v_i^2/R_i$ and the power supplied to the load amounts to $P_o = v_o^2/R_L$. The power transfer from source to load is therefore $10 \cdot \log(P_o/P_i) = 10 \cdot \log[(v_o/v_i)^2 R_i/R_L]$ dB and it depends on the load resistance and on the square of the voltage transfer. The power transfer of a characteristic system is always equal to the square of the voltage transfer, because $R_i = R_L$.

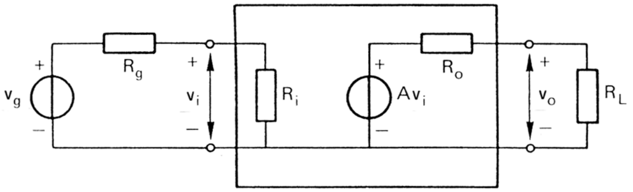


Figure 5.10. The power transfer of this system is proportional to the square of the voltage transfer.

The logarithmic representation of a voltage transfer is $20 \cdot \log(v_o/v_i)$ dB. This definition does not account for the interfacing of the system.

Example 5.3
The resistance values of the system given in Figure 5.10 are: $R_i = 100\ k\Omega$, $R_o = 0\ \Omega$, $R_L = 1\ k\Omega$ and the amplifier gain is $A = 100$.
The power transfer appears to be equal to $10\cdot\log(A^2R_i/R_L) = 10\cdot\log 10^6 = 60$ dB.
The voltage transfer of the system itself is $v_o/v_i = A = 100$ or: $20\cdot\log A = 40$ dB. This does not represent the voltage transfer from source to load which is v_o/v_g.

5.2 Signal models

Any measurement system is influenced by interference signals which obscure the measurement signals. Interference either originates from the system itself or enters from outside. Undoubtedly the designer will make every effort to minimize noise, to stop the cross talk of auxiliary signals to the measurement signal and to prevent the induction of spurious signals from outside, for instance by providing proper shielding. Ultimately it is not only the designer but also the user of the system who must be careful to keep out unwanted signals.

It remains difficult, despite careful designing, to guarantee that operations will be completely interference free. Internally generated interference (like noise) and the sensitivity to external interference should be specified by the manufacturer of the measurement system.

Interference signals introduce measurement errors. Such errors can be categorized in three ways as: destructive errors, multiplicative errors and additive errors. Destructive errors lead to the total malfunctioning of the system which is a situation that is, of course, not acceptable. Multiplicative and additive errors are permissible but only up to a specified level. Multiplicative errors result from system transfer deviations. The output of the system can be expressed as $x_o = H(1 + \varepsilon)x_i$ in which H is the nominal transfer. The output remains zero for zero input, but the transfer deviates from the specified value. The relative measurement error is ε and the absolute error, $H\varepsilon x_i$, depends on the input signal.

Multiplicative errors are caused by, for instance, component drift values (due to temperature and aging) or non-linearity.

The remainder of this section will be devoted to additive errors, in particular noise.

5.2.1 Additive errors

Additive errors become evident when the output signal differs from zero at zero input:

$$x_o = Hx_i + x_n \tag{5.12}$$

where H is the transfer of the system and x_n the error signal. Such an error signal accounts for all kinds of additive interference, like (internally generated) offset, cross talk of internal auxiliary signals or unwanted external signals entering into the system and getting mixed up with the measurement signal.

If x_n is a DC signal we say that it is offset (see Chapter 1.2):

$$x_o = Hx_i + x_{o,off} = H\left(x_i + x_{i,off}\right) \tag{5.13}$$

where $x_{o,off}$ and $x_{i,off}$ are the output and input offset. The absolute error in the output signal does not depend on the input signal (as with multiplicative errors) and the relative error increases as the input signal decreases. The offset errors are therefore always expressed in absolute terms.

The influence of all additive error signals can be represented by just two additional system input signal sources (Figure 5.11). These sources are called the equivalent error signal sources. The system itself is supposed to be error-free with all system errors concentrated into the two equivalent error signal sources, v_n and i_n.

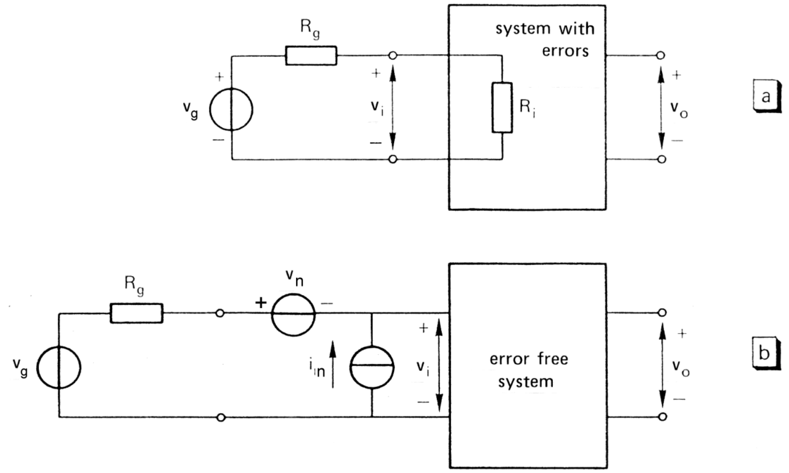


Figure 5.11. (a) A system with internal error sources; (b) modelling of the error signals by two external sources: the system is made error free.

In Figure 5.11, the signal source is modeled on a Thévenin circuit consisting of voltage source v_g and source resistance R_g. At open input terminals, the input signal of the system is $v_i = i_n R_i$, so the output signal is $v_o = H i_n R_i$. At short-circuited terminals, the output is $v_o = H v_n$. The total output signal is therefore:

$$v_o = H\left(v_i + v_n + i_n R_g\right)\frac{R_i}{R_g + R_i} \tag{5.14}$$

The output signal depends on the source resistance R_g. When the measurement system is connected to a voltage source (with a low value of R_g), v_n is the dominant output error. A voltage measurement system should therefore have a low value of v_n. In the case of current measurements, the system should have a low value of i_n to minimize the contribution made by $i_n R_g$ to the output error signal.

The error sources v_n and i_n can represent DC signals (for instance offset and drift) or AC signals (for instance noise). In the first case, the signal is expressed in terms of its momentary value while in the second case rms values are preferred. It is important to make the following distinction: the momentary value of two (or more) voltage sources in series or current sources in parallel equals the sum of the individual values. The rms value does not adhere to this linear summing rule. According to the definition (Section 2.1.2), the rms value of two voltages $v_1(t)$ and $v_2(t)$ in series is

$$v_{s,rms} = \sqrt{\frac{1}{T}\int (v_1 + v_2)^2 dt} = \sqrt{\frac{1}{T}\int \left(v_1^2 + 2v_1v_2 + v_2^2\right)dt} \tag{5.15}$$

If the two signals are completely independent (uncorrelated), the average of v_1v_2 is zero, hence

$$v_{s,rms} = \sqrt{v_{1,rms}^2 + v_{2,rms}^2} \tag{5.16}$$

This square summing rule is applied when several stochastically independent error signals are combined to form one set of equivalent error sources v_n and i_n at the system input point.

Example 5.4

Two systems, I and II, each have the equivalent error voltage source v_{n1} and v_{n2} and the equivalent error current source i_{n1} and i_{n2} (Figure 5.12a). We shall endeavor to find the equivalent set of error sources v_{ns} and i_{ns} for the two systems in series (Figure 5.12b).

Both the models given in Figure 5.12 represent the same system, so they should be completely equivalent. This equivalence allows us to calculate the error sources v_n and i_n.

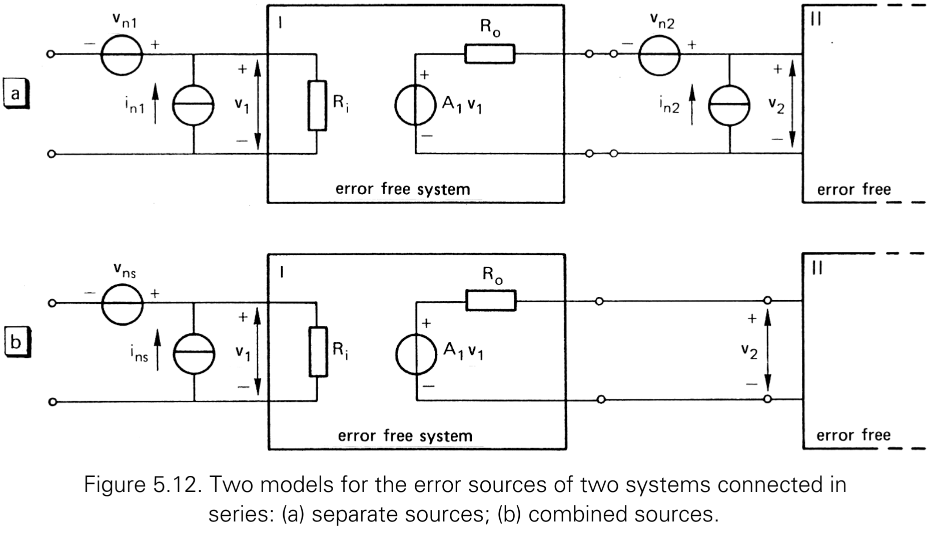


Figure 5.12. Two models for the error sources of two systems connected in series: (a) separate sources; (b) combined sources.

We take v_2 *of system II as a criterion for equivalence: this voltage must be the same in both models for any source resistance value. Take, for example, two extreme cases: $R_g = \infty$ and $R_g = 0$ (open input terminals and short-circuited input terminals) and determine for both situations the output voltage* v_2.

At open input terminals:

$$v_2 = i_{n1}R_iA_1 + i_{n2}R_o + v_{n2} = i_{ns}R_iA_1$$

so

$$i_{ns} = i_{n2} + \frac{v_{n2} + i_{n2}R_o}{A_1R_i}$$

At short-circuited input terminals:

$$v_2 = v_{n1}A_1 + i_{n2}R_o + v_{n2} = v_{ns}A_1$$

so

$$v_{ns} = i_{n1} + \frac{v_{n2} + i_{n2}R_o}{A_1}$$

This calculation only has validity for the momentary error signal values. If the calculation pertains to stochastic, uncorrelated error sources then we must write:

$$i_{ns}^2 = i_{n2}^2 + \frac{v_{n2}^2 + i_{n2}^2R_o^2}{A_1^2R_i^2}$$

and

$$v_{ns}^2 = v_{n1}^2 + \frac{v_{n2}^2 + i_{n2}^2R_o^2}{A_1^2}$$

From this example it follows that high gain in the first system is the most favorable situation with respect to total system error. The errors emanating from the second system only contribute a little to the overall error effect. In other words, the additive error of a system is almost entirely determined by just the input components if these components produce high signal gain relative to the other system parts. In a proper design, the system gain will come from its input components.

5.2.2 Noise

Any conductor or resistor exhibits thermal noise which comes from thermal movements in the material's electrons. The spectral power density (Section 2.1.3) amounts to $4kT$ (W/Hz) and is virtually independent of frequency. Thermal noise behaves like white noise.

The noise power dissipated in a resistor with resistance R is $P = V^2/R = I^2R$. Thermal noise can therefore be represented by a voltage source in series with the (noise free) resistance and has a value of $v_n = \sqrt{4kTR}$, or a current source that is in parallel to the noise-free resistance and has a value of $i_n = \sqrt{4kT/R}$ (Figure 5.13).

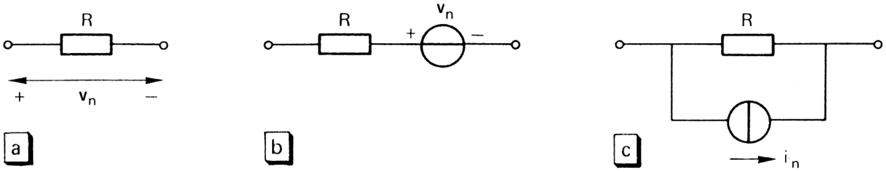


Figure 5.13. (a) Any resistor generates noise; (b) the resistance noise is represented by a voltage source or (c) a current source.

Example 5.5
To extend the measurement range of a voltmeter to include higher voltages a voltage divider is connected to the input terminal (see Figure 5.14a). We want to know what are the equivalent error sources for the whole system (Figure 5.14b) so that we can estimate how the divider contributes to system noise. Take then the amplifier's input voltage, $v_{i,}$ *as a criterion for equivalence. The equivalent current source is found by calculating* v_i *at the system's short-circuited terminals:*

$$v_i^2 = \left(\frac{R_2}{R_1 + R_2}\right)^2 v_1^2 + \left(i_2 + i_n\right)^2 \left(\frac{R_1 R_2}{R_1 + R_2}\right)^2 + v_n^2 = \left(\frac{R_2}{R_1 + R_2}\right)^2 v_s^2$$

from which follows:

$$v_s^2 = v_1^2 \left(\frac{R_1 + R_2}{R_2}\right)^2 v_n^2 + R_1^2 \left(i_2 + i_n\right)^2$$

Similarly, the equivalent current source is calculated at open inputs:

$$i_s^2 = i_2^2 + i_n^2 + \left(\frac{v_n}{R_2}\right)^2$$

These two equations clearly show how the voltage divider contributes to the total noise error in relation to the noise of the system itself.

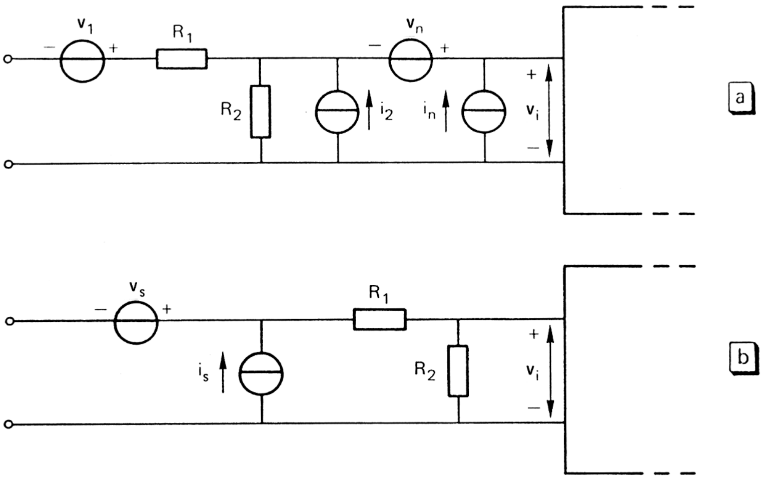


Figure 5.14. Modelling noise signals of a system with voltage divider (a) by separate error sources and (b) by combined error sources.

SUMMARY

System models

- Any two-terminal system is characterized by voltage source V_o in series with source impedance Z_g (Thévenin's theorem), or current source I_k that is in parallel to source impedance Z_g (Norton's theorem). V_o and I_k are the respective open voltage and short-circuiting current. The source resistance is $Z_g = V_o/I_k$.

- The input impedance of a system is the ratio between the (complex) input voltage and input current. The output impedance is the ratio between the output voltage and the output current. Generally, input impedance depends on load impedance and output impedance depends on source impedance.
- Linear two-port systems can be characterized using models that have two sources (a voltage source and/or a current source) and two impedances. Two system equations are used to mathematically describe the model.
- A voltage measuring instrument requires high input impedance Z_i. If the source impedance Z_g satisfies the inequality $Z_g << Z_i$, then the relative measurement error will equal about $-Z_g/Z_i$.
- A current measuring instrument requires low input impedance Z_i. If source impedance Z_g satisfies the inequality $Z_g >> Z_i$, then the relative measurement error will equal about $-Z_i/Z_g$.
- Characteristic matching is achieved when output source impedance equals input load resistance. The power transfer amounts to just $\frac{1}{2}$.
- The power transfer P_o/P_i expressed in decibels (dB) is $10{\cdot}\log(P_o/P_i)$. The logarithmic voltage transfer is defined as $20{\cdot}\log(V_o/V_i)$.

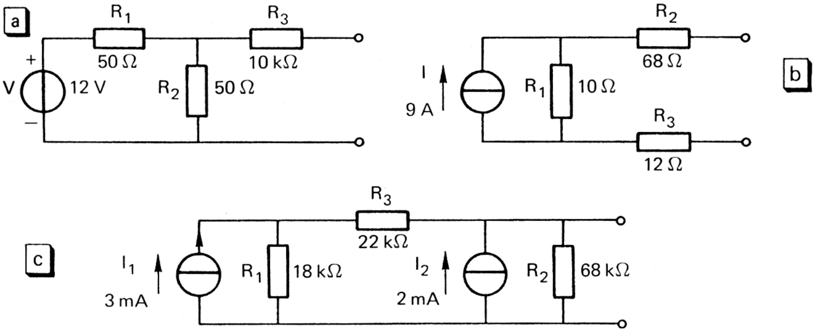
Signal models

- System errors fall into three categories: destructive, multiplicative and additive.
- Additive error signals originating from the system itself can be represented by two independent sources at the system input point, an equivalent error voltage source and an equivalent error current source.
- The rms value of the sum of two uncorrelated signals is equal to the root of the summed squares: $v_{s,rms} = \sqrt{v_{1,rms}^2 + v_{2,rms}^2 + \cdots + v_{n,rms}^2}$
- The spectral power of thermal noise is $4kT$ W/Hz. This noise can be represented by a voltage source in series with noise-free resistance, $\sqrt{4kTR}$ [V/$\sqrt{\text{Hz}}$] or by a parallel current source with a strength of $\sqrt{4kT/R}$ [A/$\sqrt{\text{Hz}}$].
- The equivalent voltage error source of a system is found by having equal output at short-circuited input terminals. The equivalent error current source is found by having equal output at open input terminals.

EXERCISES

System models

5.1 Find the Thévenin equivalent circuits for the source circuits a-c given below.



5.2 In a list of specifications, the maximum values for the rejection ratios are given as: *CMRR* (at DC): 90 dB, *CMRR* (at 1 kHz): 80 dB and supply voltage rejection ratio *SVRR*: 110 dB. Specify these quantities in µV/V.

5.3 The current from a current source with source resistance 100 kΩ is measured using a ampere meter. The maximum possible error is 0.5%. What is the requirement for the input resistance of the ammeter?

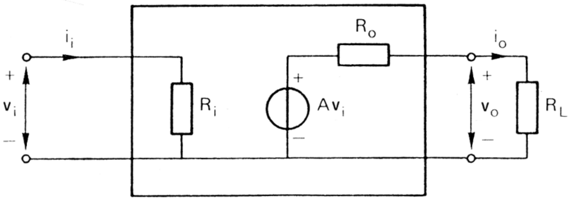
5.4 The source resistance of an unknown voltage source is measured as follows: first its voltage is measured with a voltmeter that has an input resistance of $R_i \geq 10$ MΩ, then the voltage is measured while the source is loaded with 10 kΩ. The successive measurement results are 9.6 V and 8 V.

a. Calculate the source resistance.

b. What is the relative error in the first measurement caused by the R_i load?

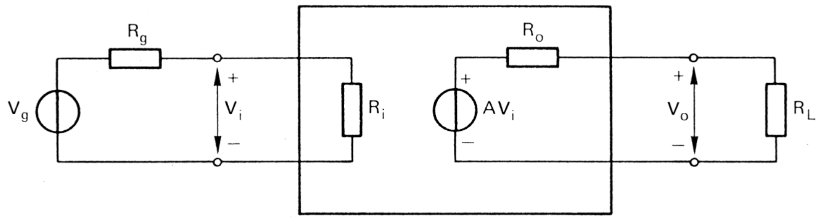
5.5 What is characteristic matching? What is characteristic impedance? How much is the power transfer at characteristic coupling?

5.6 The figure below contains a model of a voltage amplifier. Find the voltage transfer $A_v = v_o/v_i$, the current transfer i_o/i_i and the power transfer $A_p = P_o/P_i$.



Signal models

5.7 The figure below presents a voltage amplifier model. R_L is a resistor with thermal noise voltage v_{nL}, all other components are error-free. Find the equivalent error sources at the system input terminals that only derive from the noise from R_L.



5.8 Refer to the model given in the previous exercise. This time only R_i and R_o will generate noise and the respective noise contributions will be i_{ni} and v_{no}. All the other components are noise free. Determine the equivalent noise sources derived from these two resistances.

5.9 Refer once again to the previous figure. This time the offset voltage and offset current are represented by equivalent sources at the input points, V_n and I_n. At 20 °C, $V_n = 1$ mV ± 10 μV/K; $I_n = 10$ nA, the value doubles each time the temperature is raised by 10 °C. Furthermore, $R_g = 10$ kΩ, $R_i = \infty$, $R_L = \infty$ and $A = 10$.

Calculate the maximum output offset voltage:

a. for $T = 20$ °C

b. for a temperature range of $0 < T < 50$ °C.

5.10 The input power of a system is the product of its input current and its input voltage. The signal-to-noise ratio S/N is the ratio between the input signal power P_s and the input noise power P_n. Prove that S/N is independent of input resistance R_i.

6 Frequency diagrams

The complex transfer function of an electronic system usually depends on the signal frequency. A common way of visualizing the frequency dependence of the transfer function is by drawing a frequency diagram. In that way, both the modulus (amplitude transfer) and the argument (phase transfer) are plotted against the frequency. This set of features is what we call the Bode plot of the system and that is what will be discussed in the first part of this chapter. There are other ways of visualizing the frequency dependence of a complex transfer function, one of which is by means of the polar plot representing the transfer in the complex plane. This method will be discussed in the second part of the chapter.

6.1 Bode plots

6.1.1 First order systems

A Bode plot comprises two diagrams: the amplitude characteristic and the phase characteristic. Both plots have a logarithmic frequency scale. The modulus (amplitude) scale is also logarithmic and the phase (argument) is plotted along a linear scale. Modulus and argument can be rather complicated functions of frequency but despite this the Bode plot is quite easy to draw using certain approximations. The method will be explained on the basis of a simple network with one resistance and one capacitance (Figure 6.1).

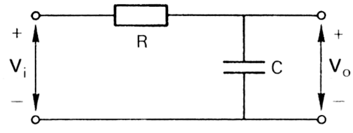


Figure 6.1. A first-order RC-network.

The complex transfer function of this network is $H(\omega) = 1/(1 + j\omega\tau)$, and $\tau = RC$, is the time constant of the network. The modulus of the transfer function equals

$$|H(\omega)| = \frac{1}{\sqrt{1+\omega^2\tau^2}} \tag{6.1}$$

To plot this function against frequency we must consider three particular conditions: very low frequencies ($\omega << 1/\tau$), very high frequencies ($\omega >> 1/\tau$) and $\omega = 1/\tau$.
In the case of low frequencies (relative to $1/\tau$) the modulus is about 1 (or 0 dB). There is no signal attenuation because the impedance of the capacitance at very low frequencies is high. With high frequencies (relative to $1/\tau$) the modulus $|H|$ can be approximated by $1/\omega\tau$. Here the modulus is inversely proportional to the frequency. With logarithmic scales, the characteristic is a descending straight line. Figure 6.2a shows these two parts, or asymptotes, of the amplitude characteristic. The asymptotes intersect the point at $\omega = 1/\tau$, the corner frequency of the characteristic. The amplitude characteristic can be approximated surprisingly well on the basis of these two asymptotes: $|H| = 1$ for $\omega\tau < 1$ and $|H| = 1/\omega\tau$ for $\omega\tau > 1$.

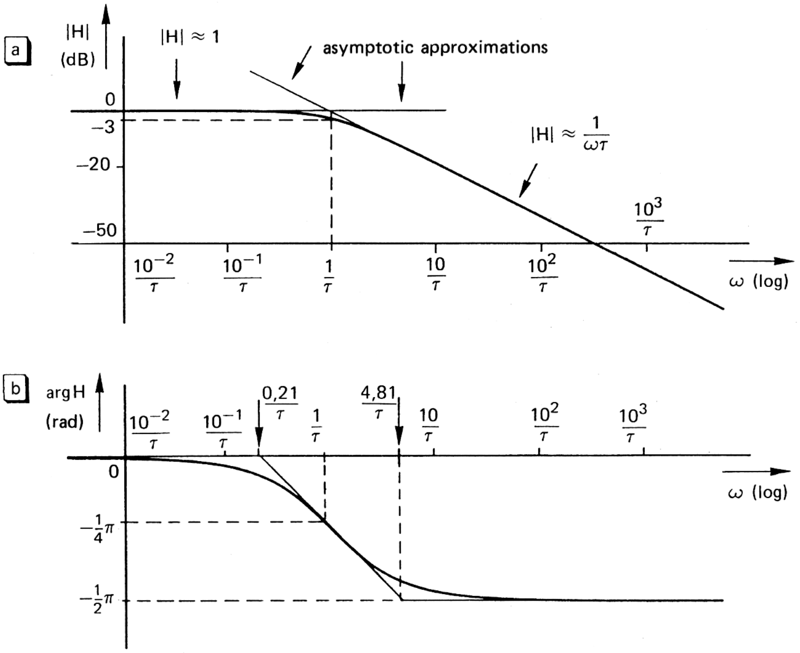


Figure 6.2. The Bode plot of the RC network of figure 6.1;
(a) amplitude transfer; (b) phase transfer.

The value of $|H|$ for $\omega\tau = 1$ equals $1/\sqrt{2}$ or $\frac{1}{2}\sqrt{2}$. Using decibel notation, this is equal to:

$$20\log|H(\omega = 1/\tau)| = 20\log 2^{-1/2} = -10\log 2 = -10 \times 0.30103 \approx -3\text{ dB}$$

This explains why the frequency where the asymptotes intersect is also sometimes called the −3dB frequency. It then becomes easy to sketch the real characteristic along the asymptotes and through the −3dB point. The deviation from the asymptotic approximation is only 1 dB for frequencies of $\omega = 1/2\tau$ and $\omega = 2/\tau$.

In the high frequency range, the modulus of the transfer is inversely proportional to the frequency: the transfer halves when the frequency is doubled (one octave). The slope of the falling asymptote is a factor 2 decrease per octave or, using decibel notation, −6dB/octave. Likewise, the transfer drops by factor 10 when there is a tenfold increase in the frequency (decade); the slope of the characteristic represents a factor 10 decrease per decade or −20dB/decade.

The phase characteristic can be approximated in a similar way, by splitting the frequency range into two parts. At low frequencies $\arg H$ approaches 0, and for high frequencies it approaches the value $-\pi/2$. These lines are two asymptotes of the phase characteristic. For $\omega\tau = 1$ the phase equals $-\pi/4$, which appears to be the point of inflection for the actual characteristic. The course of the curve can then be roughly sketched. A better approximation can be achieved when we know the direction of the tangent in the point of inflection. Its slope is found by differentiating $\arg H$ to $\log\omega$ (the vertical scale is linear, the frequency scale is logarithmic). The result is $-\frac{1}{2}\ln 10$, and it is used to find the intersections of that tangent with the asymptotes $\arg H = 0$ and $\arg H = -\pi/2$. Those points are $\omega = 0.21/\tau$ and $\omega = 4.81/\tau$ or about a factor 5 to the left and right of the point of inflection (Figure 6.2b). The actual phase transfer for $\omega\tau = 5$ and $\omega\tau = 1/5$ deviates only $(1/15)\pi$ rad from the approximations 0 and $-\pi/2$ at these points.

6.1.2 Higher order systems

We shall commence by giving transfer functions that can be written as the product of first order functions. The Bode plot of such a function is achieved by simply adding together the plots of the separate first order functions.

Example 6.1

The transfer function $H(\omega) = \dfrac{j\omega\tau_1}{1 + j\omega(\tau_2 + \tau_3) - \omega^2\tau_2\tau_3}$ *can be written as a product of the functions* $H_1(\omega) = \dfrac{j\omega\tau_1}{1 + j\omega\tau_2}$ *and* $H_2(\omega) = \dfrac{1}{1 + j\omega\tau_3}$

The function H_2 has already been discussed. The characteristic of H_1 can be approximated in a similar way: for $\omega\tau_2 << 1$, $|H_1|$ can be approximated by $\omega\tau_1$, and for $\omega\tau_2 >> 1$ the modulus of H_1 is about $\omega\tau_1/\sqrt{(\omega^2\tau_2^2)} = \tau_1/\tau_2$. Figure 6.3a shows the two individual first order characteristics $|H_1|$ and $|H_2|$ with −3dB points at $\omega = 1/\tau_2$ and $\omega=1/\tau_3$, respectively. It is assumed that $\tau_1 < \tau_2 < \tau_3$. To find the characteristic of $|H|$, we must remember that H is the product of both first order functions and that the modulus scale is logarithmic. The characteristic of $|H|$ is thus found by adding the characteristics of the two individual first order characteristics. As can be seen from Figure 6.3a, within the frequency range $1/\tau_3 < \omega < 1/\tau_2$, $|H_2|$ decreases 6 dB/octave and $|H_1|$ increases by the same amount. The total transfer therefore remains constant in this interval.

Similarly, the phase characteristic is found by adding the two individual phase characteristics of the first order transfer functions. Notice that the phase scale is *linear and that the phase of a series system is the sum of the phases of each single system.*

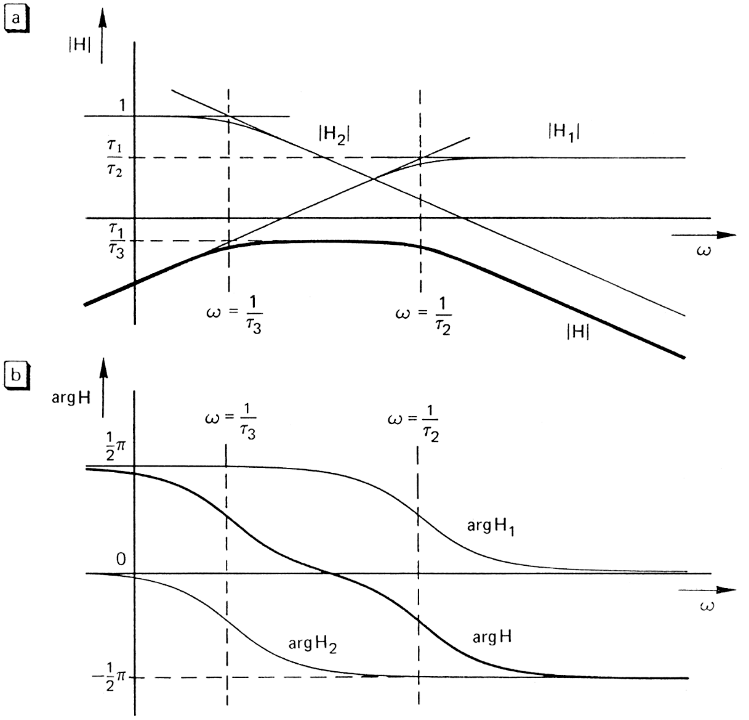


Figure 6.3. Bode plot of the transfer function $H = H_1 H_2$. It is composed of the first-order Bode plots of H_1 and H_2;
(a) amplitude characteristic; (b) phase characteristic.

This summing method offers a quick insight into the amplitude and phase characteristics of even fairly complex systems. If factorization into first order functions is not possible, this method cannot be implemented.

The next step is to study the Bode plot of second order systems, represented by the expressions

$$H(\omega) = \frac{1}{1 + j\omega\tau_1 - \omega^2\tau_2^2} \tag{6.2}$$

or

$$H(\omega) = \frac{1}{1 + 2j\omega z/\omega_0 - \omega^2/\omega_0^2} \tag{6.3}$$

We will adopt the second equation with the parameters z and ω_0. Only when $z < 0$, can the function be factorized into two first order transfer functions. Figure 6.4 shows the Bode plot for several values of $z < 1$.

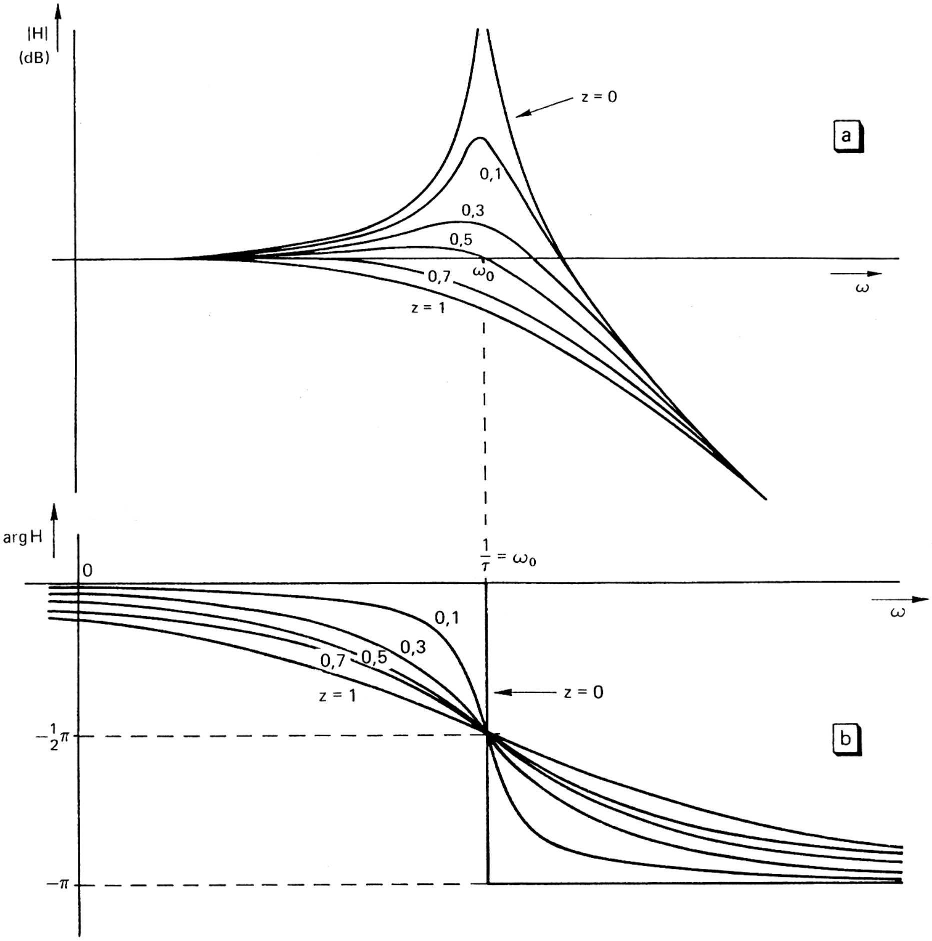


Figure 6.4. Bode plot of a second order system, for various values of the damping ratio z; (a) amplitude characteristic; (b) phase characteristic.

The transfer shows overshoot for small values of the parameter z. By differentiating $|H|$ it can easily be proved that overshoot occurs at $z < \frac{1}{2}\sqrt{2}$. This overshoot increases when z decreases. The parameter z is therefore known as the system's damping ratio. It is related to the parameter Q, the quality or relative bandwidth of the system: $Q = 1/2z$ (see Chapter 13).

When one further analyses the second order transfer function the following construction rules are obtained. At frequencies much lower than ω_0 the modulus is approximated by the line $|H| = 1$, and at frequencies much higher than ω_0 by the line $|H| = \omega_0^2/\omega^2$. The slope of the latter is −12dB/octave or −40dB/decade which means that it is twice as steep as in a first order system. The transfer shows a maximum at frequency $\omega = \omega_0\sqrt{(1 - 2z^2)}$ which amounts to $1/2z\sqrt{(1 - z^2)}$ ($z < \frac{1}{2}\sqrt{2}$). The frequency where the transfer has its peak value is called the resonance frequency. For very low z values the resonance frequency is about ω_0; that is why ω_0 is called the undamped angular frequency.

The phase characteristic has the following properties. At very low frequencies arg$H \to 0$ and at very high frequencies arg $H \to -\pi$. At $\omega = \omega_0$, the phase transfer is just $-\pi/2$, irrespective of the parameter z. The slope of the tangent through that point (the point of inflection) is found by differentiating arg H to the variable logω. The result is $-(1/z)\ln 10$. This tangent intersects the horizontal asymptotes of the phase characteristic at frequencies that are a factor $10^{\pi z/2\ln 10} = 4{,}81^z$ on both sides of ω_0.

6.2 Polar plots

The polar plot of a complex transfer function represents its modulus and argument in the complex plane where frequency ω is a parameter. The way in which polar plots are worked out will be illustrated by means of a series of examples.

6.2.1 First order functions

The first example is the RC network shown in Figure 6.1 where $H = 1/(1 + j\omega\tau)$. For each value of ω, H is represented in the complex plane. There are two ways of doing this: either by calculating $|H|$ and arg$H$ (as done in the Bode plot) or by calculating Re H and Im H. We will use the latter method.

The real and imaginary part of an arbitrary complex polynome can easily be found by multiplying the numerator and denominator by the conjugate of the denominator (the conjugate of a complex variable $a + jb$ is $a - jb$). Hence:

$$H(\omega) = \frac{1}{1 + j\omega\tau} = \frac{1}{1 + j\omega\tau} \cdot \frac{1 - j\omega\tau}{1 - j\omega\tau} = \frac{1 - j\omega\tau}{1 + \omega^2\tau^2} \tag{6.4}$$

The denominator of the resulting expression is always real so the real and imaginary part of H can be directly obtained:

$$\mathrm{Re}H = \frac{1}{1 + \omega^2\tau^2};\ \mathrm{Im}H = \frac{-\omega\tau}{1 + \omega^2\tau^2} \tag{6.5}$$

For $\omega = 0$ the real part is 1 and the imaginary part is 0. For $\omega \to \infty$ Re H and Im H are both zero. In the special case of $\omega = 1/\tau$, the real and imaginary parts equal Re $H = 1/\sqrt{2}$ and Im $H = -\pi/4$. The polar plot is depicted in Figure 6.5.

The Bode plot can be derived from the polar plot and vice versa. The polar plot in Figure 6.5 appears to be semicircular. Let Re $H = x$ and Im $= y$. If one eliminates ω from the expressions for Im H and Re H this results in:

$$x^2 + y^2 = x \tag{6.6}$$

or

$$\left(x - \tfrac{1}{2}\right)^2 + y^2 = \left(\tfrac{1}{2}\right)^2 \tag{6.7}$$

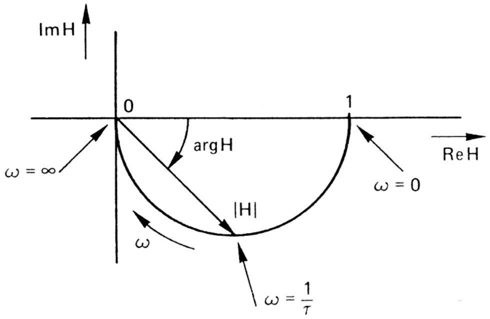


Figure 6.5. The polar plot of the function $H = 1/(1+j\omega\tau)$; for $\omega\tau = 1$, $|H| = \frac{1}{2}\sqrt{2}$ and $\arg H = -\pi/4$.

This is just the equation for a circle with center $(\frac{1}{2},0)$ and radius $\frac{1}{2}$. The polar plot only amounts to half of this circle because we only consider positive frequencies. It can be shown that the polar plot of any function $(a + jb\omega\tau)/(c + jd\omega\tau)$ is a (semi)circle; if c or d are zero, the circle degenerates into a straight line. The parameter ω moves along the plot in a clockwise direction.

We can sketch the polar plot of many simple networks without calculating the real and imaginary parts. This will be illustrated in the following examples (Figure 6.6).

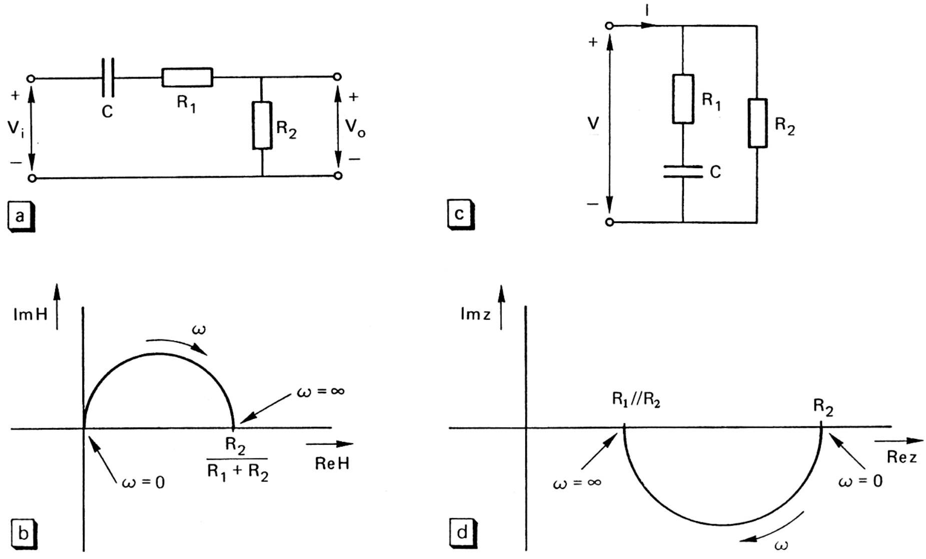


Figure 6.6. (a) A network with complex transfer function $H = V_o/V_i$; (b) the polar plot of H; (c) a network with complex impedance $Z = V/I$; (d) the polar plot of Z.

Consider the network given in Figure 6.6a. For $\omega = 0$ the transfer is zero (the capacitance blocks the signal transfer). The polar plot therefore starts at the point of origin. For $\omega \to \infty$ the transfer approaches the real value of $R_2/(R_1 + R_2)$ (the capacitance acts as a short circuit for signals); this is the end point of the semicircle, which can then be sketched directly (continually bearing in mind the clockwise direction).

Finally, we shall discuss the polar plot of the complex impedance of Figure 6.6c. The plot starts at $\omega = 0$, where Z has the real value R_2 (C has infinite impedance) and ends at $\omega \to \infty$, where $Z = R_1 // R_2$ (// means in parallel). The polar plot is a semicircle as depicted in Figure 6.6d.

6.2.2 Higher order functions

There is no simple way of constructing the polar plot for an arbitrary order two or higher complex function; the summing method used for Bode plots is of little practical use. Nevertheless, there are some rules relating to the starting and finishing points of the plot and to the direction of the tangents at those points. Some of these rules will be given below.

Consider the general transfer function:

$$F(j\omega) = \frac{a_t (j\omega)^t + a_{t+1}(j\omega)^{t+1} + \cdots + a_{T-n}(j\omega)^{T-n} + a_T (j\omega)^T}{b_n (j\omega)^n + b_{n+1}(j\omega)^{n+1} + \cdots + b_{N-1}(j\omega)^{N-1} + b_N (j\omega)^N} \tag{6.8}$$

where t and T are the lowest and highest power of $j\omega$ in the numerator and n and N the lowest and highest power of $j\omega$ in the denominator. None of the coefficients a_t, a_T, b_n and b_N are zero. Furthermore, with any physical system described by $F(j\omega)$ the order of the numerator never exceeds that of the denominator: $T \leq N$.

The starting point of the polar plot (at $\omega = 0$) is determined from

$$\lim_{\omega \to 0} F(j\omega) = \lim_{\omega \to 0} (a_t / b_n)(j\omega)^{t-n} \tag{6.9}$$

which is the approximation of $F(j\omega)$ to the lowest numerator and denominator powers. Depending upon what t and n are, this approximation will be 0 ($t > n$), a_t/b_t ($t = n$) or ∞ ($t < n$). The phase at the starting point equals $\arg(j\omega)^{t-n} = (t - n)(\pi/2)$, in other words the starting point lies either on the real axis or on the imaginary axis.

Similarly, the finishing point is found from

$$\lim_{\omega \to \infty} F(j\omega) = \lim_{\omega \to \infty} (a_T / b_N)(j\omega)^{T-N} \tag{6.10}$$

so it either amounts to a_T/b_T ($T = N$) or to 0 ($T < N$). The phase at the finishing point is $\arg(j\omega)^{T-N} = (T - N)(\pi/2)$, in other words, it either lies on the real axis or on the imaginary axis, as does the starting point.

The tangents at the starting point and the finishing point are derived respectively from $\lim_{\omega \to 0} (dF/d\omega)$ and $\lim_{\omega \to \infty} (dF/d\omega)$. The values of these limits depend on the coefficients of a and b in $F(j\omega)$, but they are always a multiple of $\pi/2$. This means that the tangents have either horizontal or vertical directions. Table 6.1 gives some results without providing further proof. These results are illustrated in the examples given below.

Table 6.1. Starting and end points of a polar plot and the direction of the tangents in these points. a) $\frac{1}{2}\pi$ sign $(a_{t+1}b_n - a_t b_{n+1})$; b) $\frac{1}{2}\pi$ sign $(a_T b_{N-1} - a_{T-1} b_N)$

	Starting point			End point	
	$t > n$	$t = n$	$t < n$	$T = N$	$T < N$
Value	0	a_t/b_n	∞	a_T/b_N	0
Tangent	$(t - n)\pi/2$	$\pm\pi/2^{a}$	$\pi + (t - n)\pi/2$	$\pm\pi/2^{b}$	$\pi + (T - N)\pi/2$

Example 6.2

The function $F(j\omega)$ is defined as:

$$F(j\omega) = \frac{j\omega\tau_1}{1 + j\omega\tau_2 + (j\omega\tau_3)^2}$$

starting point: $t = 1$; $n = 0$, so 0;
finishing point: $T = 1$; $N = 2$, so 0;
direction at the starting point: $(1 - 0)(\pi/2) = \pi/2$;
direction at the finishing point: $\pi + (1 - 2)(\pi/2) = \pi/2$.
The polar plot is depicted in Figure 6.7a. The intersection with the real axis is found from $\mathrm{Im}F = 0$, which results in $\omega = 1/\tau_3$ and $\mathrm{Re}F = \tau_1/\tau_2$.

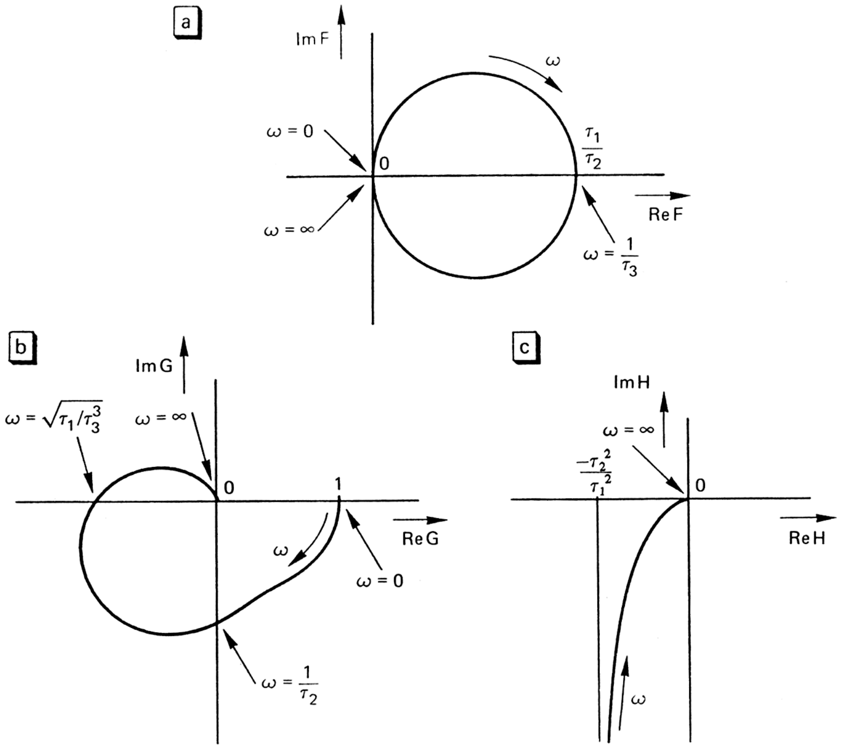

a
Im F
ω
ω = 0
0
$\frac{\tau_1}{\tau_2}$
Re F
ω = ∞
ω = $\frac{1}{\tau_3}$
b
Im G
ω = $\sqrt{\tau_1/\tau_3^3}$
ω = ∞
0
1
Re G
ω
ω = 0
ω = $\frac{1}{\tau_2}$
c
Im H
ω = ∞
$\frac{-\tau_2^2}{\tau_1^2}$
0
Re H
ω

Figure 6.7. (a) polar plot of $F=j\omega\tau_1(1 + j\omega\tau_2 + (j\omega\tau_3)^2)$
(b) polar plot of $G = 1/(1 + j\omega\tau_1 + (j\omega\tau_2)^2 + (j\omega\tau_3)^3)$;
(c) polar plot of $H = 1/(j\omega\tau_1 + (j\omega\tau_2)^2)$.

Example 6.3

The function G(jω) is defined as:

$$G(j\omega) = \frac{1}{1 + j\omega\tau_1 + (j\omega\tau_2)^2 + (j\omega\tau_3)^3}$$

starting point: $t = n = 0$, *hence* $a_t/b_n = a_0/b_0 = 1$;
finishing point: $T = 0$; $N = 3$, *hence 0;*
direction at the starting point: $\text{sign}(0 \cdot 0 - 1 \cdot 1)(\pi/2) = -\pi/2$;
direction at the finishing point: $\pi + (0 - 3)(\pi/2) = -\pi/2$.
*Figure 6.7b shows the polar plot. The intersections with the real and imaginary axes are found from Im*G = 0 *and Re*G = 0, *which are respectively:* $G(\omega = 1/\tau_2) = -j/(\tau_1/\tau_2 - \tau_3^3/\tau_2^3)$ *and* $G(\sqrt{\tau_1/\tau_3^3}) = 1/(1 - \tau_1\tau_2^2/\tau_3^3)$

Example 6.4

The function H(jω) is defined as:

$$H(j\omega) = \frac{1}{j\omega\tau_1 + (j\omega\tau_2)^2}$$

starting point: $t = 0$; $n = 0$, *so* ∞;
finishing point: $T = 0$; $N = 2$, *so 0*
direction at the starting point: $\pi + (0 - 1)(\pi/2) = \pi/2$;
direction at the finishing point: $\pi + (0 - 2)(\pi/2) = 0$.
A more precise plot will follow from further analysis done numerically (for instance, with a computer program) or analytically:

$$\text{Re}H = -\frac{\tau_2^2/\tau_1^2}{1 + \omega^2\tau_2^4/\tau_1^2};\ \text{Im}H = -\frac{1/\omega\tau_1}{1 + \omega^2\tau_2^4/\tau_1^2}$$

The real part is always negative, running from $-\tau_2^2/\tau_1^2$ $(\omega = 0)$ *to* 0 $(\omega \to \infty)$; *the domain of the imaginary part extends over all the negative values. Under these conditions it becomes possible to accurately draw the polar plot.*

It is not easy to produce a polar plot for more complicated functions. The Bode plot is easier to draw. A polar plot has the advantage of being able to represent the frequency dependence of the transfer in a single diagram.

SUMMARY

Bode plots

- The Bode plot represents the modulus and the argument for complex transfer functions versus frequency. The frequency and the modulus are plotted on logarithmic scales.
- The Bode plot of a first order function can easily be drawn with the help of asymptotic approximations for $\omega >> 1$ and $\omega << 1$.
- Point $\omega\tau = 1$ on the Bode plot of a first order system is the −3dB frequency. The deviation from the asymptotic approximation amounts to −3 dB.

- The Bode plot of a function consisting of the product of first order functions is the sum of the individual characteristics of such first order functions.
- The denominator of a second order polynome is given as $1 + 2j\omega z/\omega_0 - \omega^2/\omega_0^2$. In this expression ω_0 is the undamped frequency and z is the damping ratio. For $z < 1/\sqrt{2}$ the amplitude characteristic displays overshoot; its peak value increases as z decreases.
- The frequency of the maximum value of the amplitude transfer is called the system's resonance frequency.

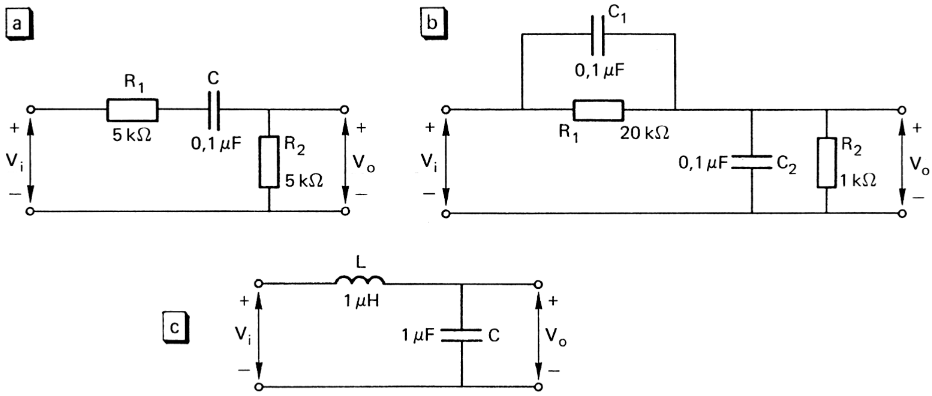
Polar plots

- A polar plot represents a complex function in the complex plane where the frequency is a parameter.
- The polar plot of a first order system describes a circle or semicircle in the complex plane. In certain cases the circle degenerates into a straight line.
- The starting point and finishing point of a polar plot are found by taking the limit of the transfer function for $\omega = 0$ and $\omega \to \infty$.
- The direction of the tangent at the starting point is $\frac{1}{2}k_1\pi$ where k_1 is the difference between the lowest power of $j\omega$ in the numerator and in the denominator of the transfer function ($k_1 > 0$).
- The direction of the tangent at the finishing point is $\pi + \frac{1}{2}k_2\pi$ where k_2 is the difference between the highest power of $j\omega$ in the numerator and the denominator ($k_2 < 0$).
- When the lowest or highest powers of $j\omega$ in the numerator and denominator of the transfer function are equal, the tangents at the starting point and the finishing point are vertically oriented.

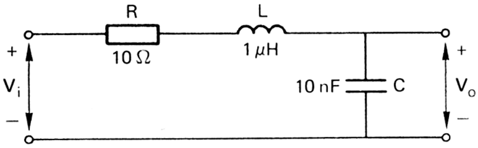
EXERCISES

Bode plots

6.1 Make a sketch of the Bode plot (an asymptotic approximation) for networks a-c in the figure below.


a
R1
C
5 kΩ
0,1 μF
R2
5 kΩ
+
Vi
−
+
Vo
−
b
C1
0,1 μF
R1
20 kΩ
0,1 μF
C2
R2
1 kΩ
+
Vi
−
+
Vo
−
c
L
1 μH
1 μF
C
+
Vi
−
+
Vo
−

6.2 Find the next parameters for the network given in the figure below: undamped frequency $\omega_{0,}$ damping ratio z, frequency for maximal amplitude transfer, peak value of the amplitude transfer and modulus of the transfer at $\omega = \omega_0$.


R
10 Ω
L
1 μH
10 nF
C
+
Vi
−
+
Vo
−

6.3 The complex transfer function of what is known as an "all pass" network is given as:

$$H(j\omega) = \frac{1 - j\omega\tau}{1 + j\omega\tau}$$

Find $|H|$ and arg$H$. Draw the Bode plot for H.

6.4 Make a sketch of the Bode plot for the following function in which $\tau = 10^{-3}$ s:

$$H(j\omega) = \frac{1}{(1 + j\omega\tau)^3}$$

6.5 Find the Bode plot of the transfer function:

$$H(j\omega) = \frac{1 + j\omega\tau_1}{(1 + j\omega\tau_2) + (1 + j\omega\tau_3)}$$

where $\tau_1 = 0.1$ s, $\tau_2 = 0.01$ s and $\tau_3 = 0.001$ s.

Polar plots

6.6 Draw the polar plots of networks a-c given in Exercise 6.1.

6.7 Draw the polar plot of the transfer function given in Exercise 6.3. Explain the term “all-pass network”.

6.8 Draw the polar plot for the following function:

$$H(j\omega)=\frac{1+j\omega\tau_1}{1+j\omega\tau_2}$$

in three cases: $\tau_1 = 2\tau_2$; $\tau_1 = \tau_2$ and $\tau_1 = \tau_2/2$

7 Passive electronic components

Up until now we have only considered models (the network elements) of electronic systems and components. In this chapter we shall be looking at the actual electronic components and discussing their properties. Electronic components are the basic constituents of electronic circuits and systems. They may be divided into passive and active components. In order to operate, active components require external power. They make signal power amplification possible. Passive components do not require auxiliary energy to operate properly. They cannot generate signal power gain, in fact it is quite the reverse: signal processing is accompanied by power loss.

In this chapter we shall be looking at passive components such as resistors, capacitors, inductors and transformers. The first part of the chapter will describe these components as electronic circuit elements. The second part of the chapter will then deal with special ways of realizing their applications as sensors.

7.1 Passive circuit components

7.1.1 Resistors

The ability of material to transport charge carriers is termed the conductivity σ of that material. The reciprocal of conductivity is resistivity ρ. The latter is defined as the ratio between the electric field strength E (V/m) and the resultant current density J (A/m^2), hence:

$$E = \rho \cdot J \tag{7.1}$$

and

$$\sigma = \frac{1}{\rho} \tag{7.2}$$

The conductivity of common materials ranges almost from zero almost up to infinity. It is determined by the concentration of charge carriers and their specific mobility within the material. On the basis of their conductivity, materials can be classified as isolators, semiconductors or conductors (Figure 7.1). Conductivity is a material property that is independent of dimensions (except in special cases as with very thin films).

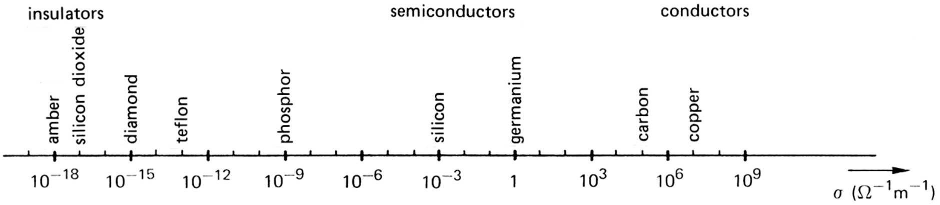


Figure 7.1 The conductivity of some materials.

The resistance R (unit ohm, Ω) of a piece of material is defined as the ratio between the voltage V across two points of the material and the resulting current I (Ohm's law). The resistance depends on the resistivity of the material as well as on its dimensions. The reciprocal of the resistance is the conductance G (unit S^{-1}, sometimes called siemens, S).

A huge range of resistor types is now commercially available. They are categorized as *fixed resistors*, *adjustable resistors* and *variable resistors*. Important criteria to consider when working with resistors are: the resistance value or range, the inaccuracy (tolerance of the value), the temperature sensitivity (temperature coefficient) and the maximum tolerated temperature, voltage, current and power. Ultimately all these properties will be determined by the material type in use and the construction. With respect to the materials, we may distinguish between carbon film, metal film and metal wire resistors.

Metal film and metal wire resistors have much better stability than the carbon film types of resistors but the latter allow for the realization of much higher resistance values. Metal wire resistors show a self-inductance and capacitance that may not always be ignored. Self-inductance can be minimized by adopting special winding methods like those implemented in expensive, highly accurate types of resistors. Usually the desired resistance value of a film resistor (made of carbon or metal) is achieved by cutting a spiral groove into a film around the cylindrical body. Table 7.1 provides an overview of the main properties of the various resistor types in common use.

Table 7.1. Some properties of different resistor types.

	Range	tolerance	temperature coefficient	maximum temperature	maximum power
carbon film	$1 \dots 10^{7}$	5 … 10	–500 … +200	155	0.2 … 1
metal film (NiCr)	$1 \dots 10^{7}$	0.1 … 2	50	175	0.2 … 1
wire wound (NiCr)	$0.1 \dots 5 \cdot 10^{4}$	5 … 10	–80 … +140	350	1 … 20

Except in very special cases, the resistance values are normalized and fit into various series (International Electro-technical Commission, 1952). The basic norm is the E-12 series where each decade is subdivided logarithmically into 12 parts. The subsequent values between 10 and 100 Ω are:

10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, 82, 100.

Resistors with narrower tolerances belong to other series, for instance the E-24 series (increasing by a factor $\sqrt[24]{10}$), E-48, E-96 and E-192. For further information on commercially available resistors the reader is referred to manufacturers' information books.

7.1.2 Capacitors

A capacitor is a set of two conductors separated from each other by an isolating material called the dielectric. The capacitance C (unit farad, F) of this set is defined as the ratio between the charge Q being displaced from one conductor to the other and the voltage V resulting between them:

$$C = \frac{Q}{V} \tag{7.3}$$

When the charge varies, so too does the voltage. The charge transport to or from the conductor per unit of time is the current I ($I = dQ/dt$), which is why the relation between the current and the voltage of an ideal capacitor is

$$I = C\frac{dV}{dt} \tag{7.4}$$

The capacitance of a capacitor is determined by the conductor and dielectric geometry. The dielectric permittivity or dielectric constant is the ability of a material to be polarized. The dielectric constant of a material that cannot form electric dipoles is, by definition, just 1. In such situations those materials behave as a vacuum.
The capacitance of a capacitor consisting of two parallel flat plates placed distance d apart which have a surface area of A and are situated in a vacuum equals

$$C = \varepsilon_0\frac{A}{d} \tag{7.5}$$

(this is an approximation which is only valid when the surface dimensions are large compared to d). In this expression, ε_0 is the *absolute* or *natural permittivity* or the permittivity of vacuum: $\varepsilon_0 = 8.85 \cdot 10^{-12}$ F/m. When the space between the conductors of a capacitor is filled with a material that has the dielectric constant ε_r, then the capacitance increases by factor ε_r:

$$C = \varepsilon_0 \varepsilon_r \frac{A}{d} \tag{7.6}$$

A large capacitance requires a dielectric material with large relative permittivity. Table 7.2 shows the relative permittivity of several materials.

Table 7.2. The dielectric constants of various materials.

material	ε_r	material	ε_r
Vacuum	1	ceramic, Al_2O_3	10
air (0°C, 1 atm)	1.000576	porcelain	6 - 8
water, at 0°C	87.74	titanate	15 - 12,000
at 20°C	80.10	plastic, PVC	3 - 5
glass, quartz	3.75	teflon	2.1
Pyrex 7740	5.00	nylon	3 - 4
Corning 8870	9.5	rubber, Hevea	2.9
mica	5 - 8	silicone	3.12 - 3.30

There is an enormous range of capacitor types. Like with resistors, capacitors are distinguished in three ways as: fixed, adjustable (trimming capacitors) and variable. Common dielectric materials are: air (for trimming capacitors), mica, ceramic materials, paper, plastic and electrolytic materials.

A capacitor never behaves as a pure capacitance; it shows essential anomalies. The leading deviations are linked to the dielectric loss, the temperature coefficient and the breakthrough voltage. We will discuss these features separately below.

When an AC voltage is applied to the terminals of a capacitor, the dipoles in the dielectric must continuously change their direction. The resulting dissipation (heat loss) is known as dielectric loss. It is modeled by a resistance in parallel to the capacitance. A quality measure for dielectric loss is the ratio between the loss resistance current and the capacitance current. This ratio is called the *loss angle* δ, defined as $\tan \delta = |I_R|/|I_C| = 1/\omega RC$ (Figure 7.2).

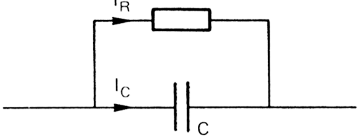


Figure 7.2. A model of a capacitor with dielectric losses.

The temperature dependence of the capacitance is caused by the expansion coefficient of the materials and the temperature sensitivity of the dielectric constant. There are capacitors with built-in temperature compensation. The non-linearity of the capacitance is generally negligible.

A very large capacitance value is achieved by having a very thin dielectric. However, electrical field strength increases as thickness decreases, thus limiting the breakthrough voltage of the capacitor.

Very high capacitance values are achieved with electrolytic capacitors. Their dielectric consists of a very thin oxide layer, formed from the material of one of the conductors (usually aluminum). The surface of the conductor is first stained which creates pores in the metal surface, thus increasing the active area. Then the surface material is anodically oxidized. The resulting aluminum oxide can withstand a high electrical field strength (of up to 700 V/μm), allowing a very thin layer to form without breakthrough occurring. In the wet aluminum capacitor the counter electrode is connected to the oxide by a layer of paper impregnated with boric acid and a second layer of (non-anodized) aluminum. In the dry type, the paper is replaced by a fibrous material (glass) impregnated with manganine. Instead of aluminum, tantalum can also be used with anodic tantalum oxide as the dielectric. Electrolytic capacitors are unipolar, in other words, they can only function correctly at the proper polarity of the voltage (as indicated on the encapsulation layer).

The highest values of the capacitance that can be obtained are 1 F (*electrolytic types*); the lowest values are around 0.5 pF and the inaccuracy is about 0.3 pF (ceramic types). Accurate and stable capacitances are realized when plastic film (for instance polystyrene) or mica are used as the dielectric. Mica capacitors have a loss angle that corresponds to $\tan\,\delta \approx 0.0002$ and are very stable: 10^{-6}/°C. They can withstand high electric field strength (60 kV/mm) and therefore also high voltage (up to 5 kV).

7.1.3 Inductors and transformers

Inductors and transformers are components based on the phenomenon of induction: a varying magnetic field produces an electric voltage in a wire surrounding that magnetic field (Figure 7.3a). Quantities that describe this effect are the magnetic induction B (unit tesla, T) and the magnetic flux Φ (unit weber, Wb). With a uniform magnetic induction field (the field has the same magnitude and direction within the space being considered), the flux equals

$$\Phi = B \cdot A \tag{7.7}$$

where A is the surface area of the loop formed by the conductor (Figure 7.3a).

In vector notation, the magnetic flux is defined as

$$\Phi = \iint_A \vec{B} \cdot d\vec{A} \tag{7.8}$$

This definition also holds for non-uniform fields. Both B and Φ are defined in such a way that the induced voltage satisfies the expression

$$V_{\text{ind}} = -\frac{d\Phi}{dt} \tag{7.9}$$

(i.e. Faraday's law of induction). The induced voltage equals the rate of change of the magnetic flux. In a uniform magnetic field the induced voltage is $v_{\text{ind}} = -AdB/dt$.

The induced voltage is not influenced when short-circuiting the wire loop. The current that will flow there will be equal to v_{ind}/R and R will be the resistance of the loop. The induced voltage increases as the rate of change of the flux (i.e. frequency) increases.

The induced voltage also increases as the loop area and number of turns increases (Figure 7.3b). Each loop undergoes the same flux change and is connected in series, so $v_{\text{ind}} = -n d\Phi/dt$, with n as the number of turns.
A varying magnetic field produces an induction voltage in a conductor. The reverse effect also exists: when current I passes through a conductor this produces a magnetic field, its strength H (unit A/m) satisfies the Biot and Savart law:

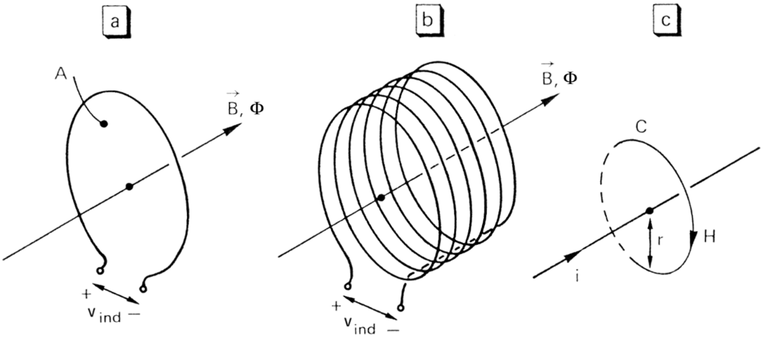


Figure 7.3. (a) A varying magnetic field produces an induction voltage in a conductor, (b) the induction voltage increases with the number of turns, (c) a current through a conducting wire generates a magnetic field.

$$H = \frac{I}{2\pi r} \tag{7.10}$$

This empiric law, which only applies to a straight wire, shows that the magnetic field strength is inversely proportional to the distance r from the wire (Figure 7.3c). Another form of this law is Ampère's law in which the integral of the magnetic field strength over a closed contour equals the total current through the contour:

$$\oint_C \vec{H} \cdot d\vec{s} = \sum_{n=1}^{N} i_n \tag{7.11}$$

This law shows that the magnetic field produced by a group of conducting wires is the sum of the fields from each individual wire.
When two conductors are mutually coupled, an electric current in one conductor produces a magnetic field which, as it varies over the course of time, induces a voltage in the other conductor. This is called mutual inductance. Inevitably, this effect also occurs in a single conductor because when a varying current goes through the conductor it will produce a varying magnetic field that induces a voltage in the same conductor. This phenomenon is called self-inductance. The direction of the induced voltage is such that it tends to reduce its origin (Lenz's law). The degree of self-inductance depends very much on the geometry of the conductor. Self-inductance increases when the conductor is coupled more closely to itself, such as inside a coil.

We have introduced two magnetic quantities: magnetic induction B and magnetic field strength H. Because of the different definitions, these two quantities do not have the same dimensions. They are joined by the equation

$$B = \mu_0 H \tag{7.12}$$

where $\mu_0 = 4\pi \cdot 10^{-7}$ Vs/Am, the *permeability* of vacuum. Obviously, the inductive coupling between two conductors (or within one conductor) depends on the properties of the medium between the conductors. The property that accounts for the ability of the material to be magnetized (which is magnetically polarized) can be described on the basis of relative permeability, μ_r. For a vacuum, the relative permeability is 1. The relation between B and H in a magnetizable medium is

$$B = \mu_0 \mu_r H \tag{7.13}$$

A high value of μ_r means a higher flux for the same magnetic field strength. For most materials μ_r is about 1. Some electronic applications require materials with a high permeability (for instance supermalloy: $\mu_r = 250{,}000$). The relative permeability of iron is about 7500.

Magnetic materials are much less ideal than dielectric materials because the magnetic losses are high and the material shows magnetic saturation, thus resulting in strong non-linearity and hysteresis.

The self-inductance L of a conductor is the ratio between magnetic flux Φ and current i through the conductor:

$$L = \frac{\Phi}{i} \tag{7.14}$$

The relation between the (AC) current through the inductor and the voltage across it is

$$v = L\frac{di}{dt} \tag{7.15}$$

We saw that the capacitance is increased by applying a high permittivity dielectric. Likewise, the self-inductance of an inductor is increased if materials that have high relative permeability are used. Inductors are constructed as coils (with or without a magnetic core) and come in various shapes (for instance solenoidal, toroidal).

Whenever possible inductors are avoided in electronic circuits as they are expensive and bulky (especially when designed for low frequencies). They display non-linearity and hysteresis (due to the magnetic properties of the core material), resistance in the wires and capacitance between the turns. Inductors are introduced to filters for signals in the medium and higher frequency ranges.

A *transformer* is based on signal transfer by mutual inductance. In its most basic form a transformer has two windings (primary and secondary windings) around a core of a material with high permeability. The ratio between the number of turns in the primary and secondary windings is the turns ratio n. For an ideal transformer with ratio $1{:}n$ turns, the voltage transfer (from primary winding to secondary winding) is just n, and the current transfer is $1/n$. A proper coupling between the two windings is achieved by overlaying them and by applying a toroidal core.

Transformers are mainly used for system power supply, for down converting mains voltage or for up converting voltages in battery-operated instruments. They are also found in special types of (de)modulators in the middle and high frequency ranges.

7.2 Sensor components

Most electric parameters of electronic components such as: resistance, capacitance, mutual inductance and self-inductance, depend on both the material properties and the dimensions of the component. This dependency allows us to construct special components that are essentially sensitive to a particular physical quantity. Such components are then suitable for use as sensors, if they meet the normal requirements with respect to sensitivity and reproducibility. Some of these sensors use the sensitivity of the material parameters to externally applied physical signals, for instance: temperature, light intensity and mechanical pressure. Others have variable dimensions in order to remain sensitive to quantities such as displacement and rotation. This section describes the basic construction and properties of sensors according to resistive, capacitive, inductive, piezoelectric and thermoelectric effects.

7.2.1 Resistive sensors

The resistance between the two terminals of a conductor equals

$$R = \rho \cdot F \tag{7.16}$$

in which ρ is the resistivity of the material and F is a geometric factor determined by the shape and dimensions of the conductor. There are two groups of resistive transducers: the first is based on variations in ρ and the second uses a variable geometry.

7.2.1.1 Temperature sensitive resistors

The resistivity of a conductive material depends on the concentration of free charge carriers and their mobility. The mobility is a parameter that accounts for the ability of charge carriers to move more or less freely throughout the atom lattice, their movement is constantly hampered by collisions. Both concentration and mobility vary according to temperature and a rate largely determined by the material.

In intrinsic (or pure) semiconductors, the electrons are bound quite strongly to their atoms, only very few have enough energy (at room temperature) to move freely. At increasing temperature more electrons will gain sufficient energy to be freed from their atom, so the concentration of free charge carriers increases as the temperature increases. Since the temperature has much less effect on the mobility of the charge carriers, the resistivity of a semiconductor decreases as the temperature increases: its resistance has a negative temperature coefficient.

In metals, all available charge carriers can move freely throughout the lattice, even at room temperature. Increasing the temperature will not affect the concentration. However, at elevated temperatures the lattice vibrations become stronger, increasing the chance of electrons colliding and hampering free movement throughout the material. The resistivity of a metal therefore increases at higher temperatures and has a positive temperature coefficient.

The temperature coefficient of the resistivity is used to construct temperature sensors. Both metals and semiconductors can be used for this, they are then called metal resistance thermometers and thermistors.

Constructing a high quality resistance thermometer requires a material (i.e. metal) with a resistivity temperature coefficient that is stable and reproducible over a wide temperature spectrum. Copper, nickel and platinum are all suitable materials. Copper is useful in the range from –140 to +120 °C and nickel from –180 to +320 °C. By far the best material, though, is platinum which has a number of favorable properties. Platinum has a high melting point (1769 °C), is chemically very stable, is resistant to oxidation and is available in very pure form. The normalized temperature range of a platinum resistance thermometer goes from -180 °C up to +540 °C. It can even reach 1000 °C but then the stability will be reduced. Platinum resistance thermometers are used as international temperature standards for temperatures varying from the boiling point of oxygen (–82.97 °C) to the melting point of antimony (+680.5 °C).

The temperature characteristic of a platinum thermometer is given as:

$$R_T = R_0(1 + \alpha T + \beta T^2 + \ldots), \tag{7.17}$$

in which R_0 is the resistance at 0 °C. The normalized value of α is $3.90802 \cdot 10^{-3}$ K^{-1} and that of β is $5.8020 \cdot 10^{-7}$ K^{-2}, according to the European DIN-IEC 751 norm. The temperature coefficient is thus almost 0.4%/K. A common value for R_0 is 100 Ω. Such a temperature sensor is called a Pt-100.

Resistive temperature sensors based on semi-conducting materials are called *thermistors.* The material of a thermistor (a contraction of the words *thermal* and *resistor*) should also have a stable and reproducible temperature coefficient. Commonly used materials are sintered oxides from the iron group (chromium, manganese, nickel, cobalt, iron). These oxides are then doped with elements of different valences to obtain lower resistivity. Several other oxides are added to improve the reproducibility. Other materials that are used for thermistors are the semiconductors germanium, silicon, gallium-arsenide and silicon-carbide.

Thermistors cover a temperature range from –100 °C to +350 °C. Their sensitivity is much larger than that of resistance thermometers. Thermistors can also be very small, which makes them highly suitable for measuring temperatures inside or on top of small objects. Compared to resistance thermometers, thermistors are less time stable and show much greater non-linearity.

As explained before, the resistance of most semiconductors has a negative temperature coefficient. Much the same goes for thermistors. That is why a thermistor is also called an NTC-thermistor or just an NTC. The temperature characteristic of an NTC satisfies the next equation:

$$R_T = R_0 \exp B\left(\frac{1}{T} - \frac{1}{T_0}\right) \tag{7.18}$$

with R_0 the resistance at T_0 (0 °C) and T the (absolute) temperature (in K). The temperature coefficient (or sensitivity) of an NTC is:

$$\alpha = \frac{1}{R}\frac{dR}{dT} = -\frac{B}{T^2}\ (\mathrm{K}^{-1}) \tag{7.19}$$

The parameter B ranges from 2000 to 5000 K, for instance, at B = 3600 K and room temperature (T = 300 K), the sensitivity amounts to –4% per K. To obtain stable sensitivity, thermistors are aged by means of special heat treatment. A typical stability value after ageing is +0.2% per year.

Alongside of NTC thermistors there are also PTC thermistors where the temperature effect differs essentially from that of a thermistor. PTCs are made up of materials from the barium-strontium-lead-titanate complex and have a positive temperature coefficient over a rather restricted temperature range. Using various materials, a total range is covered that goes from about −150 °C to +350 °C. Within the range of a positive temperature coefficient the characteristic approximates

$$R_T = R_0 e^{BT}, \quad T_1 < T < T_2 \tag{7.20}$$

The sensitivity in that range is B (K^{-1}) and can be as high as 60% per K. PTC thermistors are rarely used for temperature measurements because of their lack of reproducibility. They function mainly as safety components to prevent overheating or overload.

7.2.1.2 Light sensitive resistors

The resistivity of some materials depends on the intensity of incident light (the photoresistive effect). A resistor made up of such a material is the LDR (light-dependent resistor or photoresistor). A common material is cadmium sulfide. In the absence of light, the concentration of free charge carriers is low which is why the resistance of the LDR is high. When light falls on the material, free charge carriers are generated, concentration increases and so the resistance decreases as the intensity increases. The light sensitivity depends on the wavelength of the light and is maximal at about 680 nm (red light). Below 400 nm and above 850 nm the LDR is not usable (Figure 7.4).

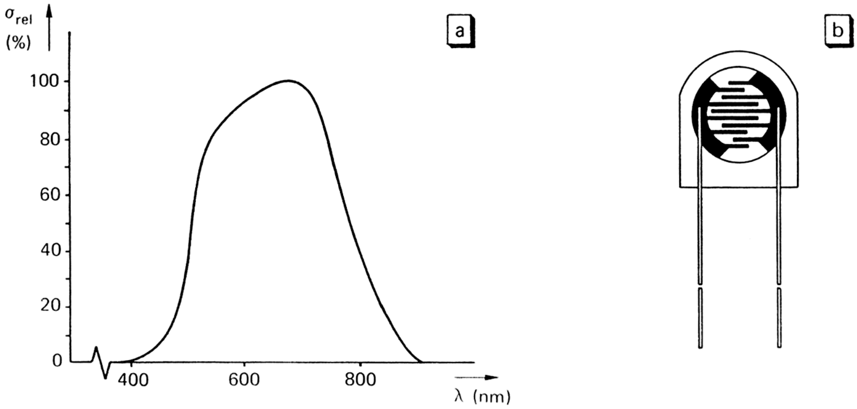


Figure 7.4. (a) The relative conductivity of an LDR versus wavelength at constant light intensity, (b) an example of an LDR.

Even in complete darkness resistance appears to be finite, this is the dark resistance of the LDR, which can be more than 10 MΩ. The light resistance is usually defined as resistance at an intensity of 1000 lux and it may vary from 30 to 300 Ω for the different types. Photoresistors change their resistance value fairly slowly. The response time from dark to light is about 10 ms while from light to dark the resistance only varies by about 200 kΩ/s.

7.2.1.3 Force sensitive resistors

When an electric conductor, like for instance a metal wire, is stressed its resistance increases because the diameter decreases and the length increases. Semiconducting materials also change their resistivity when subjected to a mechanical force (piezoresistive effect).

Strain gauges are sensors that are based on these effects. They are used to measure force, pressure, torque and (small) changes in length in, for instance, mechanical constructions. Such strain gauges consist of a meander-shaped metal wire or foil (Figure 7.5a), fixed on an isolating flexible carrier (e.g. an epoxy).

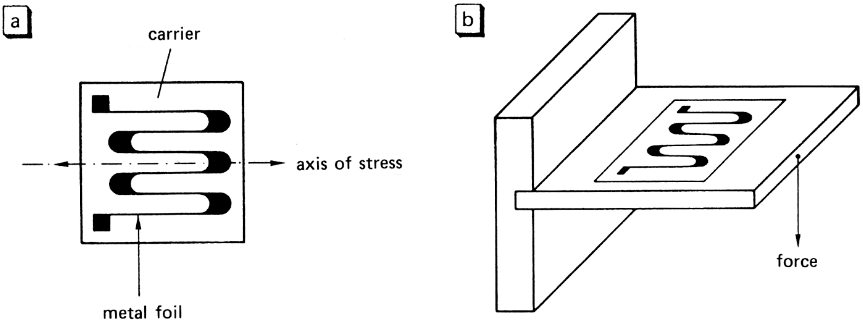


Figure 7.5. (a) An example of a strain gauge,
(b) an application of a strain gauge for the measurement of bending.

A strain gauge undergoes the same strain as the construction onto which it is glued (Figure 7.5b). The resistance change is proportional to the change in length. This proportionality factor, or sensitivity, is called the strain gauge factor or simply the gauge factor. Metal strain gauges have a gauge factor of about 2. Semiconductor strain gauges have a much higher gauge factor (up to 200) but are less stable, less linear and do have a higher temperature sensitivity. If measurement errors caused by temperature changes are to be eliminated there must be temperature compensation. This is achieved by, for instance, using a bridge circuit with two or four strain gauges.

For further information on strain gauges the reader is referred to the data sheets issued by manufacturers.

7.2.1.4 Resistive displacement sensors

A common resistive displacement sensor is the potentiometer which is a conductive track with a movable ruler (Figure 7.6). The conductor may consist of a spiralized wire, a homogeneous track of carbon or a conductive polymer. Linear displacement sensors have a straight conductive track but angular displacement sensors have a circular (one-turn) or helix-shaped track (multi-turn potentiometer).

The salient quality parameters are non-linearity (determined by the homogeneity of the conductive track), the resolution (infinite for film types) and the temperature coefficient. There are linear potentiometers ranging from several mm up to 1 m, and angular sensors having up to 10 revolutions. Non-linearity is well below 0.1%. There are also potentiometers with a prescribed, for instance logarithmic, non-linear relationship between displacement (rotation) and resistance change. Such potentiometers do directly do certain arithmetic signal processing.

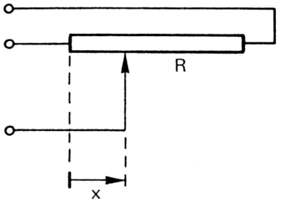


Figure 7.6. A linear resistive displacement sensor.

7.2.2 Inductive sensors

Inductive sensors are based on changes in self-inductance L or on the mutual inductance M of a component which usually consists of two parts that can move relative to each other. Figure 7.7a shows an inductor (coil) with a movable core. Its self-inductance L varies according to the position of the core. Due to leak fields caused by the winding and at the ends of the coil, the self-inductance varies only linearly according to displacement Δx over a limited range (Figure 7.7b). Another disadvantage of this type of sensor is that an impedance (self-inductance) needs to be measured which is somewhat more difficult than simply dealing with a voltage or current.

In this respect the construction of Figure 7.8a is better. This sensor, which is called a linear variable differential transformer or short LVDT consists of one primary winding, two secondary windings connected in series but wound in opposite directions and a movable core. When the core is just in the center of this symmetric construction the voltages of the secondary windings are equal but opposite, so their sum is zero. A displacement of the core from the center position results in an imbalance between the two secondary outputs. The total output amplitude v_0 thus increases according to displacement Δx. The phase of v_0 with respect to v_i is 0 or π, depending on the direction of displacement.

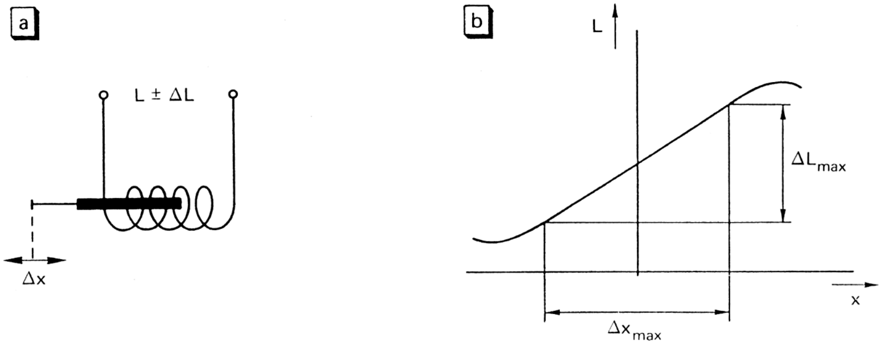


Figure 7.7. (a) A coil with a movable core as a displacement sensor, (b) self-inductance as a function of core displacement.

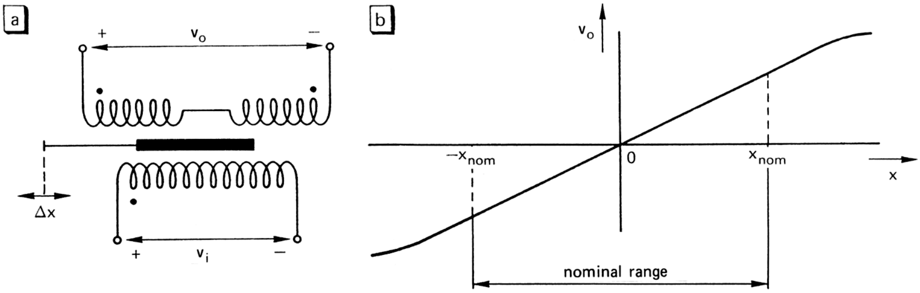


Figure 7.8. (a) A linear variable differential transformer, (b) output voltage amplitude versus displacement.

LVDTs are made for displacements that go from several mm up to 1 m with a non-linearity that is less than ±0.025% (of the nominal displacement). The sensitivity of an LVDT is expressed as mV (output voltage) per mm (displacement) per V (input voltage) and ranges from 10 to roughly 200 mV/mm·V, depending on the type.

Inductive sensors with a movable core can also be used for the measurement of a force (with a spring) or acceleration (with a mass-spring system). In fact, all displacement sensors can be used to measure velocity or acceleration by once or twice differentiating their output: $v = dx/dt$; $a = dv/dt$.

Figure 7.9a shows another type of inductive sensor. It consists of a coil with a fixed core. The coil is connected to an AC voltage source so it generates a varying magnetic field in front of the sensor head. If a conductor is close to the sensor, a voltage will be induced in the conductor (Section 7.1.3). As there is no preferred current path in the object, currents will flow in arbitrary directions, such currents are called eddy currents. Due to mutual coupling, these currents in turn induce, in their turn, a voltage in the sensor coil that opposes the original voltage and thus reduces self-inductance. The effect depends on the distance between the coil and the object and can be used for the construction of a contactless displacement sensor. Such an eddy-current displacement sensor is used for measuring (short) distances and for a number of other applications, such as measuring the number of revolutions (tachometer), Figure 7.9b.

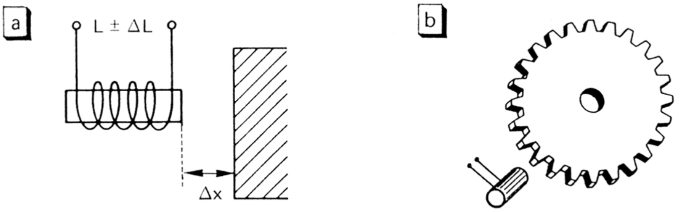


Figure 7.9. (a) Eddy-current sensor, (b) an example of an application of an eddy-current sensor as a tachometer.

7.2.3 Capacitive sensors

The capacitance of a set of conductors is given as

$$C = \varepsilon_0 \varepsilon_r F \tag{7.21}$$

where F is a factor determined by the geometry of the conductors. For a capacitor consisting of two parallel flat plates, $F = A/d$ while A is their surface area and d the distance between the plates. Capacitive sensors are based on changes in ε_r or in F. Some examples of the first possibility are:

- capacitive temperature sensors using temperature dependence ε_r. These are used for the measurement of temperatures close to absolute zero,
- capacitive level sensors. Usually a linear, tubular capacitor is connected vertically in a vessel or tank. The dielectric varies according to the level, as does the capacitance.
- capacitive concentration sensors. The dielectric constant of materials like powders or grains depends on the concentration of a particular (dielectric) substance (for instance water).

Capacitive displacement sensors are based on variations in the factor F. Figure 7.10 gives some examples of such sensors.

A balanced configuration (Figure 7.10d and e) is used where high sensitivity is required, common changes (for instance due to temperature) are cancelled out. There are cylindrical constructions (similar to the LVDT) that have extremely low non-linearity (0.01%) over the whole nominal range (± 2 mm up to ± 250 mm).

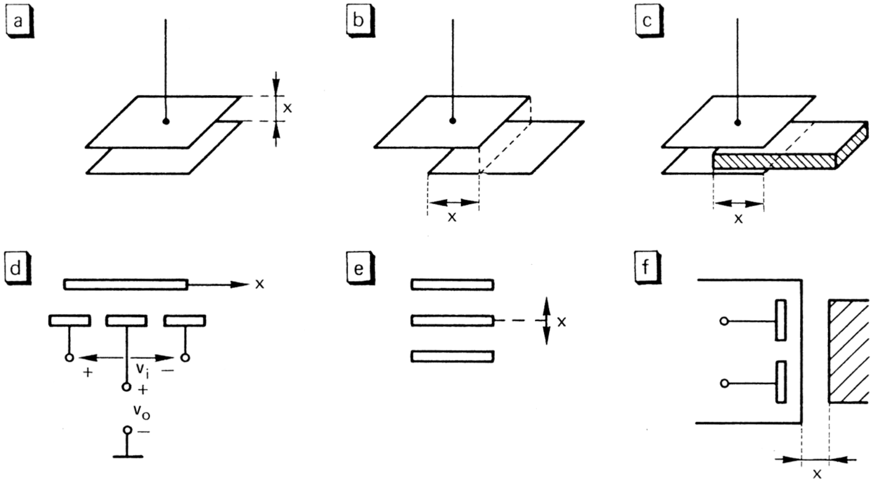


Figure 7.10. Examples of capacitive displacement sensors, (a) variable distance between the electrodes, (b) variable electrode surface area, (c) variable dielectric, (d) & (e) balanced configurations (differential capacitor) (f) as a proximity sensor.

7.2.4 Thermoelectric sensors

7.2.4.1 The Seebeck effect

The free charge carriers in different materials have different energy levels. When two different materials are connected to each other, diffusion will cause the rearrangement of the charge carriers. This will precipitate a voltage difference at that junction. Naturally, neutrality is maintained within the construction as a whole. The value of this junction potential will all depend on the type of materials used and the temperature.

In a coupling consisting of two *a* and *b* material junctions in series (Figure 7.11), two voltages are generated at both junctions but with opposite polarity. The voltage across the end points is zero, as long as the junction temperatures are equal.

If the two junctions have a different temperature the thermal voltages do not cancel, so there is a net voltage across the end points of the coupling that complies with the expression:

$$V_{ab} = \beta_1(T_1 - T_2) + \tfrac{1}{2}\beta_2(T_1 - T_2)^2 + \ldots \tag{7.22}$$

This phenomenon is called the Seebeck effect, after its discoverer, Thomas Johann Seebeck (1770-1831). V_{ab} is the Seebeck voltage. The coefficients β depend on the materials and slightly also on the temperature. The derivative of the Seebeck voltage to the variable T_1 is:

$$\frac{\partial V_{ab}}{\partial T_1} = \beta_1 + \beta_2(T_1 - T_2) + \ldots = \alpha_{ab} \tag{7.23}$$

and it is called the Seebeck coefficient. The value depends on the materials and on the temperature as well. The Seebeck coefficient can always be presented as the difference between two other coefficients: $\alpha_{ab} = \alpha_{ar} - \alpha_{br}$, with α_{ar} and α_{br} being the Seebeck coefficients of the couplings for materials *a* and *r* (*a* is the reference material) and *b* and *r*, respectively. Usually the reference material is lead but sometimes it is copper or platinum.

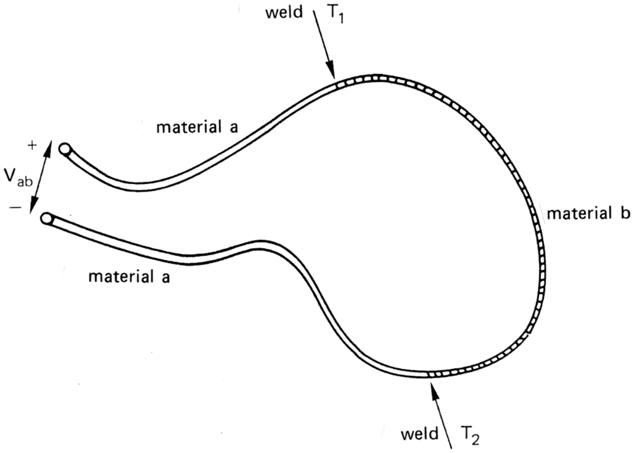


Figure 7.11. A thermocouple generates DC voltage V_{ab} between its end points if the junction temperatures are not equal.

The Seebeck effect is the basis for the thermocouple, a thermoelectric temperature sensor. One of the junctions, the reference or cold junction, is kept at a constant, known temperature (for instance 0 °C). Meanwhile the other junction (the hot junction) is connected to the object whose temperature has to be measured. Actually, the thermocouple only measures the temperature difference, not an absolute temperature. Thermocouple materials should have a Seebeck coefficient that is high (to achieve high sensitivity), a low temperature coefficient (to obtain high linearity) and be stable in terms of time (for good long-term stability of the sensor).

Thermocouples cover a temperature range from almost 0 K to over 2800 K, and belong to the most reliable and accurate temperature sensors in the 630 °C to 1063 °C range. Table 7.3 reveals the properties of various thermocouples.

The sensitivity of most couples shown in Table 7.3 depends on the temperature. For example, the sensitivity of the couple copper/constantane reaches, at 350 °C, a value of 60 μV/K.

All in all, metal thermocouples have a relatively low sensitivity. To obtain a sensor with a higher sensitivity, a number of couples are connected in series. All cold junctions are thermally connected to each other and all the hot junctions are connected to each other as well. The sensitivity of such a thermopile is *n* times that of a single junction where *n* is the number of couples.

Table 7.3. Some properties of various thermocouples.

materials	Composition	sensitivity at 0°C (μV/K)	temperature range (°C)
iron/constantane	Fe / 60% Cu + 40% Ni	45	0 - 760
copper/constantane	Cu / 60% Cu + 40% Ni	35	–100 - 370
chromel/alumel	90% Ni + 10% Cr 94% Ni + 2% Al + rests	40	0 -120
platinum/platinum+ rhodium	Pt / 90% Pt + 10% Rh	5	0 - 1500

7.2.4.2 The Peltier effect and the Thomson effect

Experimentation has shown that when electric current I flows through the junction of two materials a and b, there is heat flow Φ_w from the environment to the junction. The direction of the heat flow changes when the current is reversed. The heat flow appears to be proportional to the current I:

$$\Phi_w = -\Pi_{ab}(T)I \tag{7.24}$$

with Π the Peltier coefficient, named after its discoverer, Jean Peltier (1785-1845). The Peltier coefficient depends only on the types of materials involved and the absolute temperature. In a series with two junctions, heat flows from one junction to the other (Figure 7.12), raising the temperature at one end of the connection and lowering it at the other end. The Peltier effect can thus be applied to cooling.

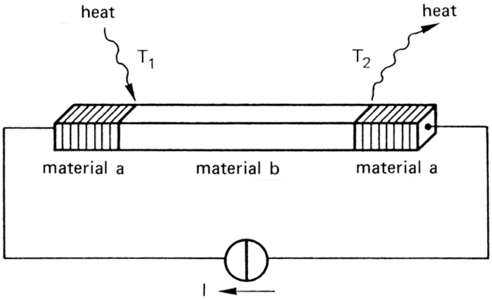


Figure 7.12. Basic construction of a Peltier element: when a current flows through a couple of junctions one of these junctions heats up and the other cools down.

A Peltier element is a cooling device based on the Peltier effect. To achieve a reasonable cooling power, many cold junctions are connected thermally in parallel, in a similar way to the thermopile. The hot junctions are connected to a heat sink to reduce the heating of the hot face and to create a lower temperature at the cold face. The Peltier effect is, after all, to generate a temperature difference. The temperature difference across a Peltier element is limited by dissipation because of the Peltier current itself. The maximum difference is determined by the Peltier coefficient, the electrical resistance, and the thermal resistance between the cold and hot faces of the element. Theoretically, the maximum temperature difference is about 65 °C (at room temperature). This value is obtained from selected materials like doped bismuth-telluride (Bi_2Te_3). Cooling down to lower temperatures is achieved by piling up the Peltier elements. However, the maximum temperature difference is far less than that of the sum of the single elements.

From the theory of thermodynamics it follows that at the junction of materials a and b, heat production is $+\Pi_{ab}(T)I$ on the one hand, and $+\alpha_{ab}(T)IT$ on the other hand. So, there is a relation between the Peltier and Seebeck coefficients:

$$\Pi_{ab}(T) = -T\alpha_{ab}(T) \tag{7.25}$$

with T being the absolute temperature at the junction.

Heat exchange with the environment also occurs when current I flows through a homogeneous conductor that has a temperature gradient. That is known as the Thomson effect, named after William Thomson (Lord Kelvin), (1824-1907). The heat flow appears to be proportional to the current I and the temperature gradient in the conductor:

$$\Phi_w = \mu(T)I\frac{dT}{dx} \tag{7.26}$$

and $\mu(T)$ is the Thomson coefficient of the material. Theoretically, μ is equal to $T(d\alpha/dT)$. Usually, in a thermocouple, all three effects: the Seebeck effect, the Peltier effect and the Thomson effect, occur simultaneously. To prevent the thermocouple from self-heating or self-cooling it must be connected to a measurement circuit with high input resistance. The current through the junctions is then kept to a minimum.

7.2.5 Piezoelectric sensors

Some materials show electrical polarization when subjected to a mechanical force, this is called the piezoelectric effect. The polarization originates from a deformation of the molecules and results in a measurable surface charge or, via the relation $Q = CV$, a voltage across the material. The relation between force and charge is extremely linear, up to several thousand volts.

The sensitivity of piezoelectric materials is expressed in terms of the piezoelectric charge constant d (C/N or m/V).

A material with natural piezoelectricity is quartz (SiO_2). Many ceramic materials and some polymers can be made piezoelectric by poling at high temperatures (above the Curie temperature). Piezoelectric materials are used for the construction of piezoelectric force sensors and accelerometers. In the latter case, a well defined mass (a seismic mass) is connected to the piezoelectric crystal of the sensor.

A piezoelectric accelerometer (Figure 7.13a) has a strong peak in the amplitude characteristic (Figure 7.13b), due to the mass spring system of the construction. Another important shortcoming of a piezoelectric sensor is its inability to measure static forces and accelerations, the charge generated by a constant force will gradually leak away via the (high) resistance of the piezoelectric crystal, or along the surface of the edges.

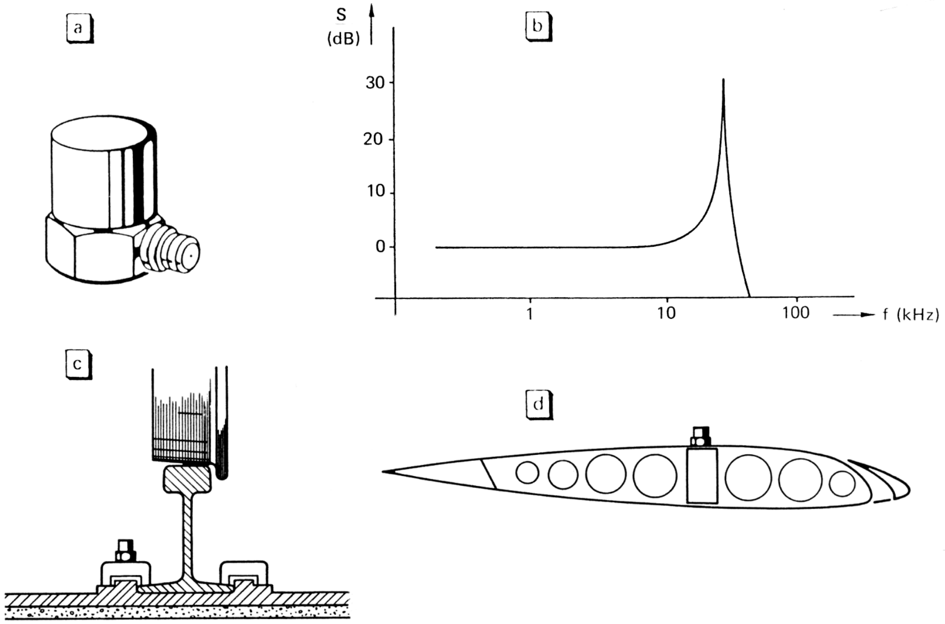


Figure 7.13. (a) An example of a piezoelectric accelerometer,
(b) the sensitivity of a piezoelectric accelerometer shows a pronounced peak in the frequency characteristic,
(c) & (d) two application examples of piezoelectric accelerometers as vibration sensors: on a rail and on the model of an airplane wing.

Table 7.4 shows the range and the sensitivity of piezoelectric sensors as force sensors and as accelerometers. The various values depend very much on the size and construction of the sensor.

Table 7.4. The range and sensitivity of piezoelectric force sensors and accelerometers.

	force sensor	accelerometer
range	10^2 - 10^6 N	$5 \cdot 10^3$ - 10^6 ms^{-2}
sensitivity	2 - 4 pC/N	0.1 - 50 pC/ms^{-2}
resonance frequency		1 - 100 kHz

The piezoelectric effect is reversible which is what makes the material suitable for the construction of actuators as well like, for instance, acoustic generators. Recently, linear motors have also been designed using the piezoelectric effect.

SUMMARY

Passive circuit components

- The resistivity or specific resistance ρ of a material is the ratio between the electric field strength E and the resulting current density J: $E = \rho J$; the reciprocal of the resistivity is the conductivity, $\sigma = 1/\rho$.
- Ohm's law for a resistor is: $V = RI$
- The capacitance of a set of two conductors is the ratio between the charge Q and the voltage V: $Q = CV$. For two parallel plates that have surface area A and are distance d away from each other the capacitance is $C = \varepsilon_0 A/d$ (in vacuum).
- The dielectric constant or relative permittivity ε_r of a material is a measure of its electric polarizability (i.e. the formation of electrical dipoles). For vacuum, $\varepsilon_r = 1$, a capacitor with a dielectric material has a capacitance that is ε_r as much as in vacuum.
- One of the drawbacks of a capacitor is the loss angle δ, defined as tan $\delta = I_R/I_C$, where I_R and I_C are the respective resistive and capacitive currents through the capacitor.
- The voltage induced by a varying magnetic field is $V_{ind} = -d\Phi/dt$. With Φ the magnetic flux is defined in turn as the integral of magnetic induction B: $\iint \overline{B} \cdot d\overline{A}$
- The magnetic field strength H provoked by a current through a straight wire at distance r is $H = I/2\pi r$.
- The magnetic permeability of a material is a measure of its magnetic polarizability (i.e. formation of magnetic dipoles) and is defined through the relation $B = \mu_0 \mu_r H$. For vacuum, $\mu_r = 1$.
- The self-inductance of a coil (inductor) is the ratio between the flux produced Φ and the current through the inductor: $\Phi = LI$.
- The voltage transfer of an ideal transformer is equal to the ratio of the numbers of primary and secondary turns, n; the current transfer amounts to $1/n$.

Sensor components

- For the measurement of physical quantities and parameters, special resistors are designed with resistances that vary in a reproducible way according to physical quantities such as: temperature, light intensity, relative displacement or angular displacement.
- Resistance thermometers are based on the positive temperature coefficient of the resistance of a metal. A common type of sensor is the Pt-100: this sensor has a resistance of 100.00 Ω at 0°C, and a sensitivity of about 0.39 Ω/K.
- Thermistors have a negative temperature coefficient: their resistance decreases exponentially as temperature increases.
- Strain gauges are resistive sensors where resistance varies when mechanical force is applied. Temperature compensation is invariably necessary.

- An LVDT (linear variable differential transformer) is a cylindrical transformer with a movable core. The voltage transfer is proportional to the linear displacement of the core.
- Capacitive sensors are based on the variation of their active dimensions or on dielectric properties. The first possibility is used for the measuring of linear and angular displacement; the second for measuring, for instance: level, the moisture content of a substance or thickness.
- Thermocouples are temperature sensors based on the Seebeck effect. With thermocouples, temperature differences can be measured accurately over a wide range.
- The Peltier effect is another thermoelectric effect that can be used for cooling.
- The piezoelectric effect is used for the construction of force sensors and accelerometers. Static measurements cannot be made. The frequency characteristic of a piezoelectric accelerometer shows a pronounced peak, due to mechanical resonance.

EXERCISES

Passive circuit components

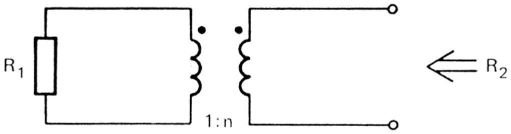
7.1 Give the relation between the current i and the voltage v of an (ideal) capacitor with capacitance C. The same question is put for the charge q and the voltage v.

7.2 What is the loss angle of a capacitor?

7.3 Give the relation between the current i and the voltage v of an (ideal) coil with self-inductance L. The same question can be put for the flux Φ and the current i.

7.4 What are the relations between magnetic induction B, magnetic flux Φ and magnetic field strength H? Give, too, the dimensions of these three quantities.

7.5 Show that the impedance R_2 between the terminals of the secondary winding of the ideal transformer is equal to n^2R_1 (see the figure below).



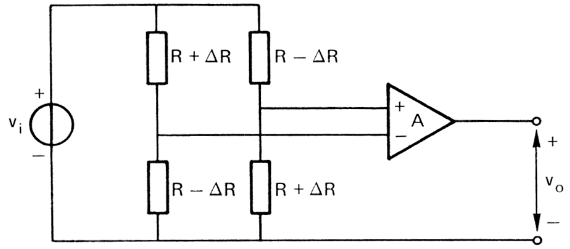
Sensor components

7.6 The transfer characteristic of a platinum resistance is given as: $R(T) = R_0(1 + \alpha T + \beta T^2)$, with $R_0 = 100\ \Omega$, $\alpha = 3.9 \cdot 10^{-3}\ \text{K}^{-1}$ and $\beta = -5.8 \cdot 10^{-7}\ \text{K}^{-2}$. Calculate the maximum non-linearity error relative to the line $R(T) = R_0(1 + \alpha T)$, in a temperature range going from −50 °C to +100 °C. Express the error in % and in °C.

7.7 Both the connecting wires of the Pt100 sensor given in Exercise 7.6 have resistance r. What is the maximum value of r if the measurement error due to this resistance is to be kept below an equivalent value of 0.1 °C?

7.8 A chromel-alumel thermocouple is used to measure the temperature of an object with a required inaccuracy of less than 0.5 °C. Find the maximum permitted input offset voltage for the voltage amplifier.

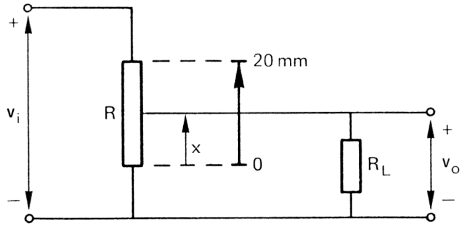
7.9 A strain gauge measurement circuit consists of four strain gauges connected to a circular bar in such a manner that the resistances vary according to torsion, as indicated in the figure below.



The required resolution of the system is a torsion corresponding to a relative resistance change of $\Delta R/R = 10^{-6}$. Find the minimum CMRR of the amplifier A.

7.10 Give the Thévenin equivalent circuit of a thermocouple, taking into account the resistance of the connecting wires. Give also the Norton equivalent of a piezoelectric force sensor, bearing in mind the capacitance and the leakage resistance of the crystal.

7.11 A linear potentiometric displacement sensor has a range from 0-20 mm and a total resistance of $R = 800\ \Omega$ (see the figure below). The sensor is connected to an ideal voltage source and loaded with a measurement instrument that has an input resistance of $R_L = 100\ \text{k}\Omega$.



What is the maximum non-linearity error and what is its displacement?

8 Passive filters

In conjunction with a particular property, usually the frequency of the signal components, an electronic filter will enable signals to separate. In this respect it can be viewed as a system with a prescribed frequency response.
Filters have a wide area of application, some examples being:

- to improve signal-to-noise ratio. If the frequency range of a measurement signal differs from that of the interference or noise signals then the latter can be removed from the measurement signal by filtering.
- to improve the dynamic properties of a control system. Circuits with a particular frequency response are connected to a control system in order to meet specific stability requirements and other criteria.
- as signal analysis instruments. Many frequency-selective measurement instruments contain special filters like spectrum analyzers and network analyzers.

In this chapter we will be discussing filters that are composed solely of passive components (resistors, capacitors, inductors). Active filters will be examined in Chapter 13.

The advantages of passive filters are:

- their high linearity (passive components are highly linear),
- their wide voltage and current range,
- the fact that no power supply is required,

The disadvantages are:

- that the required filter properties cannot always be combined with other requirements, for instance with respect to input and output impedances,
- that the inductors used for low frequency applications are rather bulky and ideal inductance or self-inductance is difficult to realize,
- that not all kinds of filter characteristics can be produced only with resistors and capacitors.

There are four main filter types (Figure 8.1): low-pass, high-pass, band-pass and band-reject or notch filters. Signal components with frequencies that lie within the pass band are transferred properly, other frequency components are attenuated as much as possible.

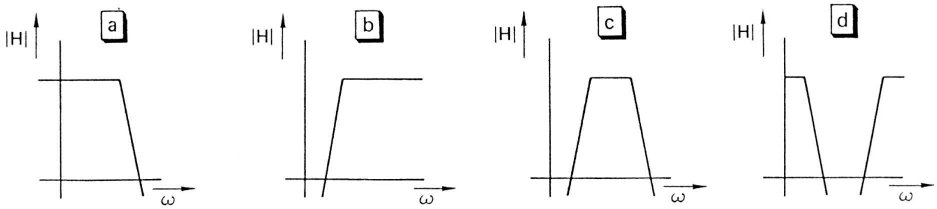


Figure 8.1. Amplitude transfer characteristic of four main filter types: (a) lowpass; (b) highpass; (c) bandpass and (d) notch filter.

There is no such thing as filters that have an ideal flat band pass frequency characteristic up to the cut-off frequency and zero transfer beyond it. The selectivity can be increased by increasing the network order (hence the number of components). The first part of this chapter deals with simple, inductorless filters going up as far as the second order. The second part deals with filters of a higher order.

8.1 First and second order RC-filters

8.1.1 Low-pass first-order RC-filter

Figure 8.2 shows the circuit diagram and the amplitude characteristic (frequency response of the amplitude) of a low-pass filter consisting of just one resistance and one capacitance. The transfer characteristic has already been discussed (Section 6.1.1). The amplitude transfer is about 1 for frequencies up to $1/2\pi\tau$ Hz ($\tau = RC$). Signals with a higher frequency are attenuated: the transfer decreases by a factor of 2 as the frequency is doubled. The cut-off frequency of this RC-filter is equal to the −3dB frequency.

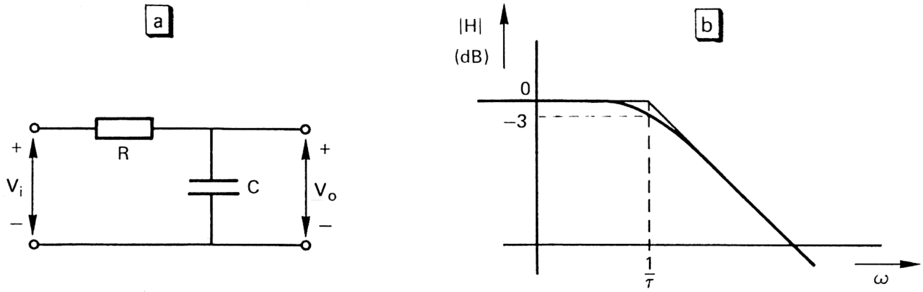


Figure 8.2. (a) A lowpass RC-filter of the first order; (b) with corresponding amplitude transfer diagram.

The frequency characteristic represents the response to sinusoidal input voltages. Let us now look at the response to another signal, that of the stepwise change of the input (Figure 8.3a). When the input voltage takes a positive step upwards the capacitor will still be uncharged, so its voltage will be zero. The current through resistance R is, at that moment, equal to $i = E/R$, where E is the height of the input step. From then on, the capacitor is charged with this current. Gradually the voltage across the capacitor increases while the charge current decreases until, finally, the charge current drops to zero. In its steady-state, the output voltage equals the input voltage E.

If this filter's step response is to be more precisely calculated the system differential equation will have to be solved (Section 4.2). The solution appears to be

$$v_o = E\left(1 - e^{-t/\tau}\right) \tag{8.1}$$

If the input voltage jumps back to zero again the discharge current will at first be equal to E/R. This current decreases gradually until the capacitance is fully discharged and the output voltage returns to zero (Figure 8.3b). The differential equation solution to this situation derived in Section (4.2):

$$v_o = Ee^{-(t-t_o)/\tau} \tag{8.2}$$

an exponentially decaying voltage. The "speed" of the filter is characterized by the parameter τ, which is called the system's time constant. The time constant of this filter appears to be the reciprocal of the −3dB frequency.

The step response can be drawn quite easily when we realize that the tangent at the starting point intersects the end value at time $t = \tau$ after the input step.

Now we can compare the system behavior as described in the time domain and in the frequency domain. For high frequencies the transfer approximates $1/j\omega\tau$ (Section 6.1). This corresponds to integration in the time domain (Section 4.2). From Figure 8.3 it can be deduced that a periodic rectangular input voltage results in a triangular-shaped output voltage, in particular for high input frequencies. This is why the RC-network is called an integrating network, even though it only has integrating properties for frequencies much higher than $1/\tau$.

Example 8.1

A sine shaped measurement signal of 1 Hz is masked by an interference signal with equal amplitude and a frequency of 16 Hz (Figure 8.4a). This composite signal is applied to the lowpass filter of Figure 8.4b in order to improve the signal-to-noise ratio. When the cut-off frequency (or −3dB frequency) is 4 Hz ($\tau = 1/8\pi$ s), the amplitude transfer at 1 Hz appears to be 0.97 (going on the theory expounded in Chapters 4 and 6), whereas the amplitude transfer at 16 Hz is only 0.24 (see Figure 8.4c).

To achieve better suppression of the interference signal at 16 Hz we may shift the −3dB frequency of the filter away from the signal frequency to, for instance, 2 Hz. If that is done, the respective amplitude transfers for 1 Hz and 16 Hz will be 0.89 and 0.12.(Figure 8.4d). It is hardly possible to discriminate better than that between the measurement signal and the interference signal because a further decrease in the −3dB frequency would reduce the measurement signal itself. For a −3dB frequency of just 1 Hz the transfers are 0.71 (or −3dB) and 0.06. The best signal-to-noise (signal-

to-interference) improvement ratio in this example is a factor of just 16, because the frequency ratio is 16 and the slope of the filter characteristic is 6 dB/octave.

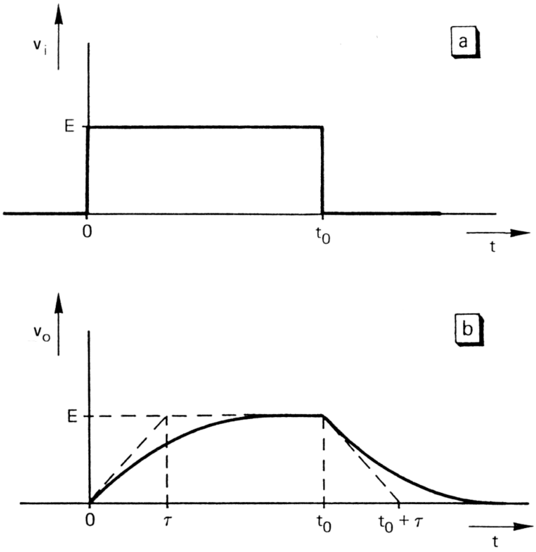


Figure 8.3. (a) Step voltage at the input of an RC lowpass filter; (b) corresponding step response.

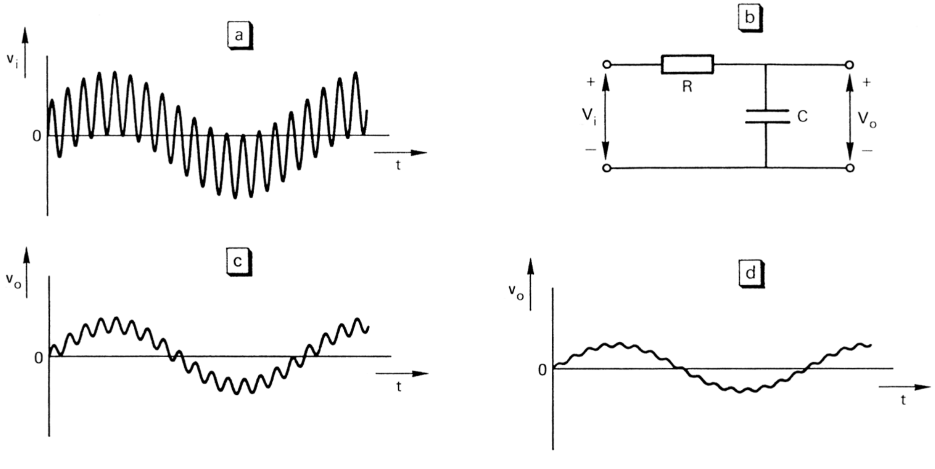


Figure 8.4. (a) The input signal v_i and (b) the filter from example 8.1; (c) the output signal v_o for $\tau = 1/8\pi$ s; (d) the output for $\tau = 1/2\pi$ s.

We will now show how non-ideal matching influences the filter properties. Suppose that the filter in Figure 8.2 is connected to a voltage source with source resistance R_g and loaded with load resistance R_L (Figure 8.5).

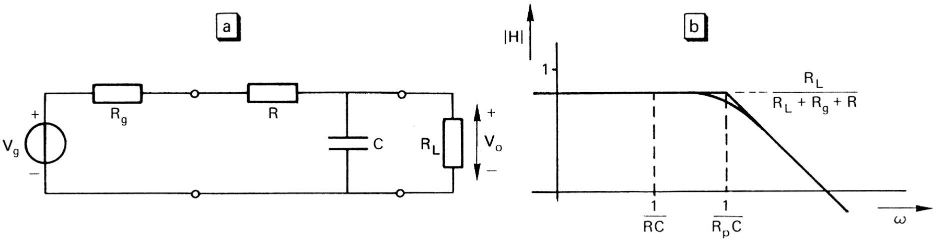


Figure 8.5. (a) A lowpass RC-filter connected to a source with source resistance R_g and loaded with a load resistance R_L; (b) the corresponding transfer characteristic.

The complex transfer function of the total circuit will then be:

$$\frac{V_o}{V_g}=\frac{R_L}{R_L+R_g+R}\cdot\frac{1}{1+j\omega R_pC} \tag{8.3}$$

with

$$R_p=\frac{R_L\left(R_g+R\right)}{R_L+R_g+R} \tag{8.4}$$

The transfer in the pass band is lowered and the cut-off frequency also depends on the load resistance. Higher source resistance and higher load resistance will move the cut-off frequency to lower frequencies. When introducing an RC-filter to a system, the source and load resistance influences must not be overlooked.

8.1.2 Highpass first-order RC-filter

Figure 8.6 shows the circuit diagram and the corresponding amplitude transfer of a first-order RC high-pass filter. This filter comprises a single resistance and capacitance. Its transfer function is

$$\frac{V_o}{V_i}=\frac{j\omega\tau}{1+j\omega\tau} \tag{8.5}$$

with $\tau=RC$. Signal components with a frequency above $1/2\tau$ are transferred unattenuated; at lower frequencies the transfer decreases by 6 dB/octave.
This filter's step response can be found in a similar way to that described in the preceding section. The exact expression for the step response can be found by solving the network's differential equation: $v_o=f(v_i)$. The solution appears to be:

$$v_o=Ee^{-t/\tau} \tag{8.6}$$

For low frequencies, the transfer approximates $j\omega\tau$, which corresponds to a differentiation in the time domain. This network is therefore called a differentiating network, although its differentiating properties are only exhibited at low frequencies ($\omega \ll 1/\tau$).

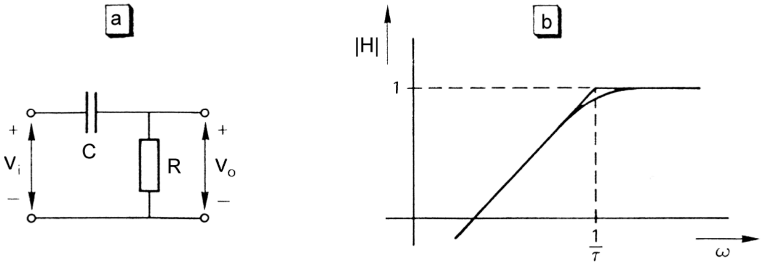


Figure 8.6. (a) A highpass RC-filter of the first order; (b) the corresponding amplitude transfer.

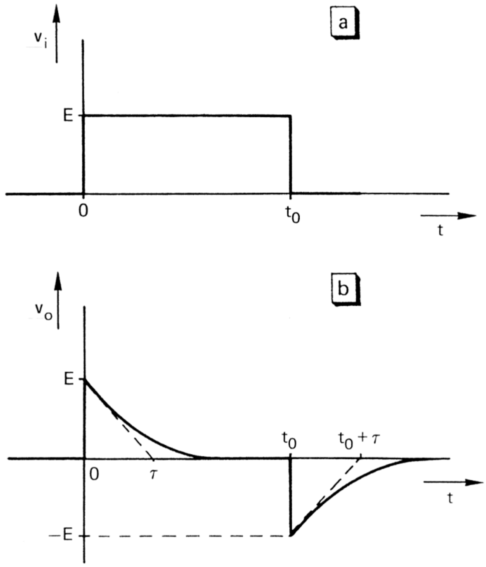


Figure 8.7. (a) A step voltage at the input of a highpass RC filter; (b) the corresponding step response.

Example 8.2

The Figure 8.4a signal (drawn again in Figure 8.8a) is now connected to the high-pass filter input in Figure 8.8b. We consider the 16 Hz signal frequency to be the measurement signal and the 1 Hz component to be the interference signal. At a cut-off frequency of 4 Hz, the amplitude transfer at 1 Hz will be about 0.24, and for a frequency of 16 Hz it will be about 0.97 (see Figure 8.8c).

To further suppress the 1 Hz interference signal, the cut-off frequency of the filter must be switched to a higher value. Suppose that the –3dB frequency is 8 Hz. In that case the amplitude transfer at 1 Hz will only be 0.12 while at 16 Hz it is almost 0.89.

For the same reason as that given in Example 8.1, it is not possible to reduce the interference signal without simultaneously reducing the measurement signal. Figure 8.8d demonstrates a favorable situation: a cut-off frequency at just 16 Hz, resulting in amplitude transfers of 0.06 at 1 Hz and 0.71 (–3dB) at 16 Hz.

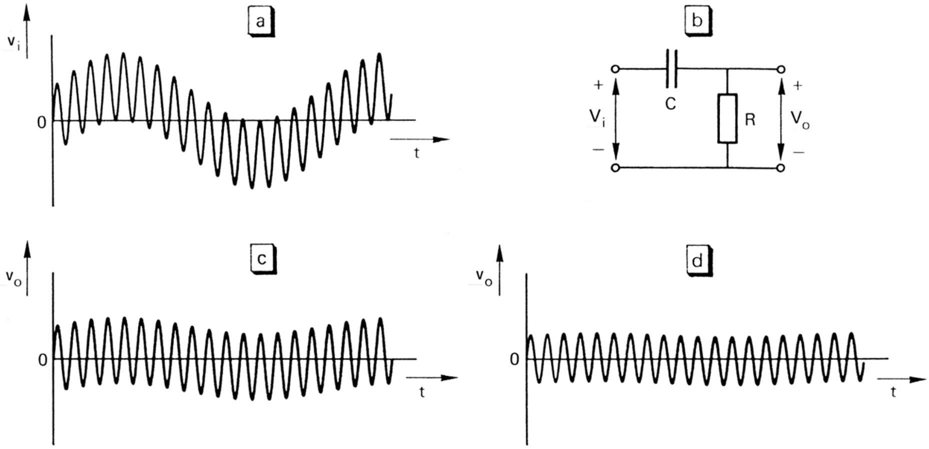


Figure 8.8. (a) The input signal v_i and (b) the filter circuit from Example 8.2; (c) the output voltage v_o at $\tau = 1/8\pi$ s; (d) the output for $\tau = 1/32\pi$ s.

The frequency response of a high-pass RC filter changes when it is connected to a source impedance and a load impedance. We shall consider a source resistance situation and a load capacitance situation (Figure 8.9).

The complex transfer of the total system is:

$$\frac{V_o}{V_i} = \frac{j\omega RC}{1 + j\omega\left(RC + R_g C + RC_L\right) - \omega^2 R_g R C_L C} \tag{8.7}$$

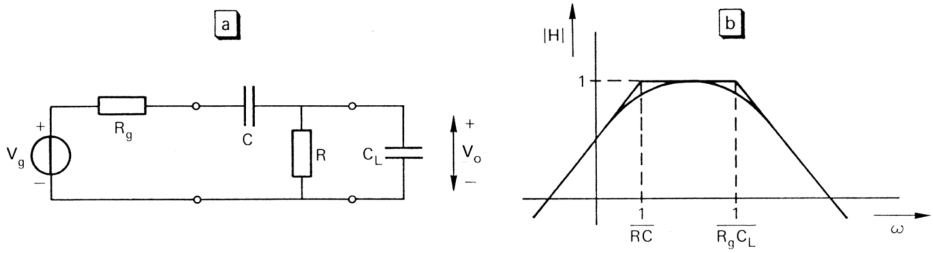


Figure 8.9. (a) A highpass RC-filter with source resistance and load capacitance; (b) due to the capacitive load, the character of the filter is changed.

The exact position of the cut-off frequencies is found by writing the denominator as $(1 + j\omega\tau_1)(1 + j\omega\tau_2)$ and solving the equations $\tau_1\tau_2 = R_g R C_L C$ and $\tau_1 + \tau_2 = RC + R_g C + RC_L$. The cut-off frequencies are $1/\tau_1$ and $1/\tau_2$ (6.1.2). Their approximated values can be found more easily by using the conditions $R_g << R$ and $C_L << C$ which are the conditions for optimal voltage matching. The denominator can then be factorized into $(1 + j\omega RC)(1 + j\omega R_g C_L)$, from which will follow the cut-off frequencies $1/RC$ (i.e. the original high-pass filter frequency) and $1/R_g C_L$ (an additional cut-off frequency derived from the source and load). This example illustrates again the importance of taking into account the source and load impedances.

8.1.3 Bandpass filters

A simple way to obtain a band-pass filter is by connecting a low-pass filter and a high-pass filter in series (Figure 8.10).

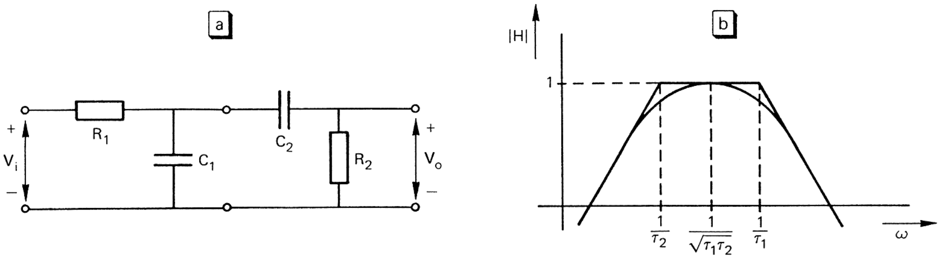


Figure 8.10. (a) A bandpass RC filter composed of a lowpass filter and a highpass filter in series; (b) the corresponding amplitude transfer characteristic.

Because of the mutual loading, the transfer is not equal to the product of the two sections individually. From calculations it follows that:

$$H = \frac{V_o}{V_i} = \frac{j\omega\tau_2}{1 + j\omega(\tau_1 + \tau_2 + a\tau_2) - \omega^2\tau_1\tau_2} \tag{8.8}$$

with $\tau_1 = R_1C_1$, $\tau_2 = R_2C_2$ and $a = R_1/R_2$. The ratio τ_2/τ_1 determines the width of the pass band (the bandwidth of the filter). The transfer has a maximum at $\omega = 1/\sqrt{\tau_1\tau_2}$, for which $|H| = 1/(1 + a + \tau_1/\tau_2)$. Because the denominator contains the term $(j\omega)^2$, the filter is said to be of the second order. The slope of the amplitude characteristic is + and −6dB/octave: its selectivity is rather poor. This filter cannot separate two closed frequencies very well.

There are different ways of creating similar band-pass filter amplitude transfers, using the same number of components, for instance by changing the section order. Figure 8.11 shows some of the possibilities.

The complex transfer of the filter in Figure 8.11a is:

$$H = \frac{V_o}{V_i} = \frac{j\omega\tau_1}{1 + j\omega(\tau_1 + \tau_2 + a\tau_2) - \omega^2\tau_1\tau_2} \tag{8.9}$$

with $\tau_1 > \tau_2$; the transfer of the circuit in Figure 8.11b is:

$$H = \frac{V_o}{V_i} = \frac{j\omega(\tau_1/a)}{1 + j\omega(\tau_1 + \tau_2 + \tau_1/a) - \omega^2\tau_1\tau_2} \tag{8.10}$$

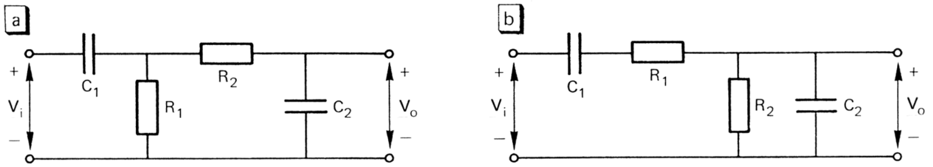


Figure 8.11. Some examples of a second-order RC-bandpass filter.

In both cases, $\tau_1 = R_1C_1$, $\tau_2 = R_2C_2$ and $a = R_1/R_2$. The filter of Figure 8.11b has the advantage that both capacitive and resistive loads can easily be combined with C_2 and R_2, respectively. They do not introduce additional cut-off frequencies to the characteristics. The same holds for a source resistance or capacitance in series with the input. Obviously the position of the cut-off frequencies may be changed but the shape of the characteristic will not change.

8.1.4 Notch filters

There are many types of notch filters composed of only resistors and capacitors. A fairly common type is the symmetric double-T filter (or bridged-T filter) that is depicted in Figure 8.12.

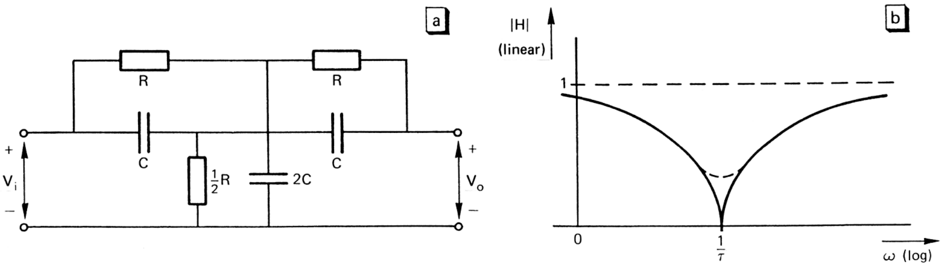


Figure 8.12. (a) The symmetric double-T notch filter; (b) the corresponding amplitude transfer characteristic.

When voltage matching is perfect (zero source resistance and infinite load resistance), the complex transfer function is:

$$\frac{V_o}{V_i} = \frac{jv}{4 + jv} \qquad (8.11)$$

with $v = \omega/\omega_o - \omega_o/\omega$ and $\omega_o = 1/RC$. This transfer is just zero for $v = 0$, or $\omega = 1/RC$. So this filter completely suppresses a signal with the frequency ω_o. This only applies, however, if the circuit component values exactly satisfy the ratios as given in the figure. Even a small deviation results in a finite transfer at $\omega = 1/RC$ (dashed curve in Figure 8.12).

8.2 Filters of higher order

The selectivity of first order filters is rather poor: the slope of the amplitude transfer characteristic is no more than 6dB/octave or 20 dB/decade. The same holds for the second order band-pass filter from the preceding section. It is very easily possible to realize second order band-pass filters with a much higher selectivity, using the principle of resonance. An example is the combination of a capacitance and an inductance. Resonance can also be achieved without inductances, using active circuits. Such filter types, which are discussed in Chapter 13, have a high selectivity but only for a single frequency, their bandwidth is very narrow.

Filter characteristics with extended flat pass bands and steep slopes are achieved by increasing the order of the networks. Band-pass filters can be composed of low-pass and high-pass filters. Low-pass and high-pass filters greatly resemble each other so this section will be restricted to low-pass RC-filters.

8.2.1 Cascading first-order RC-filters

Consider a number of n first order RC low-pass filters with equal time constants τ that are connected in series. The transfer function of such a system can be written as $H = 1/(1+j\omega\tau)^n$. It is presumed that the sections do not load each other (this effect will be discussed later). For frequencies much higher than $1/\tau$, the modulus of the transfer approximates $|H| = 1/(\omega\tau)^n$; this means that the transfer decreases by factor $2n$ when the frequency is doubled. In other words, the slope is $6n$ dB/octave. The attenuation at $\omega = 1/\tau$ amounts to −3dB for a single section, so it is $-3n$ dB for an n order filter. The attenuation within the pass band is considerable.

Example 8.3

Figure 8.13 illustrates a filter consisting of 3 low-pass RC-sections, each with a cut-off frequency of $\omega_c = 1/RC$. *The impedance of the components in the successive sections is increased by a factor of 10, starting from the source side. The effect of mutual loading can therefore be neglected and the transfer will approximate* $1/(1+j\omega\tau)^3$.

At the cut-off frequency ω_c *of this filter, the transfer is* $(\_\sqrt{2})^3 \approx 0.35$*, so attenuation is substantial. The amplitude transfer of a first order low-pass filter at a frequency of* $\omega_c/10$ *is about 0.995, so it closely approaches unity. With the third-order filter in this example, the value is* $(0.995)^3 \approx 0.985$*, which is 1.5% lower than the ideal situation. If the filter was designed with equally valued resistances and capacitances, then matters would be even worse. The transfer in that case would be:*

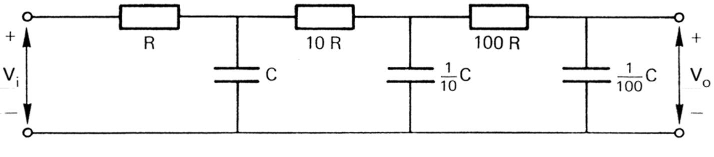


Figure 8.13. A third order lowpass filter composed of three cascaded first-order RC sections.

$$H = \frac{1}{1 + 6j\omega\tau - 5\omega^2\tau^2 - j\omega^3\tau^3} \tag{8.12}$$

its modulus for $\omega = \omega_c$ would be 0.156, and for $\omega = \omega_c/10$ it would only be 0.89 (see Figure 8.14, curve e).

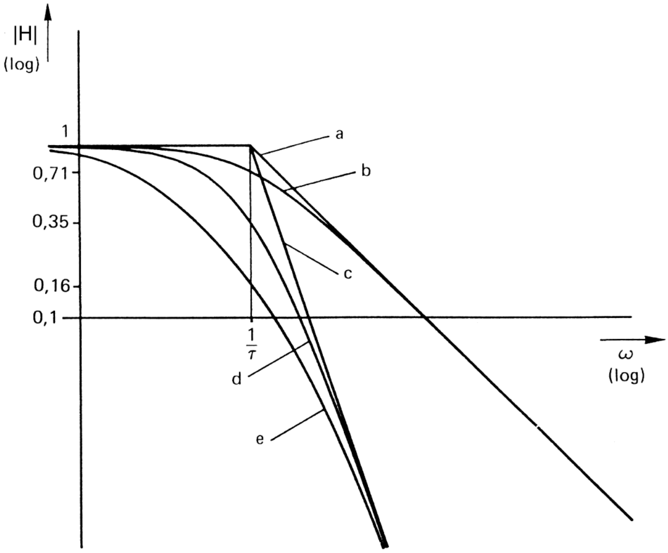


Figure 8.14. The amplitude transfer characteristic of (a) the asymptotic approximation of a first order lowpass filter; (b) a real first-order filter; (c) the asymptotic approximation of a third order filter; (d) the filter from Figure 8.13; (e) the same filter, now with equal resistances and equal capacitances.

8.2.2 Approximations of the ideal characteristics

What clearly emerges from Example 8.3 is that the coefficients of the terms $j\omega$, $(j\omega)^2$ and $(j\omega)^3$ have a great influence on the shape of the transfer characteristic. One could try to find the right coefficients for the best approximation of the ideal characteristic for a given filter order. In order to find those optimal values several criteria may be used leading to different characteristics. Three such approximations will be discussed here.

- *Butterworth filter*

The criterion is a maximally flat characteristic in the pass band that reaches the desired cut-off frequency $\omega = \omega_c$. The shape of the characteristic beyond the cut-off frequency will not be considered. The transfer function of order n, which satisfies this criterion, appears to be:

$$H = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}} \tag{8.13}$$

In terms of poles and zeroes (Section 6.2), the poles of the complex transfer function are positioned equidistantly on the unity circle in the complex plane.

- *Chebyshev filter*

The criterion is a maximally steep slope from the cut-off frequency. The shape in the pass band will not be considered here. It appears that in satisfying the condition for the slope, the pass band shows a number of oscillations. This number increases as the order increases. To find the proper values of the filter components, the designer of a Chebychev filter usually makes use of special tables.

- *Bessel filter*

The Bessel filter has an optimized step response. The criterion is a linear phase transfer up to cut-off frequency. The step response shows no overshoot. Here, too, the filter designer uses tables to find the proper component values.

Figure 8.15 draws a qualitative comparison between the amplitude transfer characteristics and the step responses made by the three types.

None of the filter types discussed here can be realized just with resistances and capacitances. One needs to use either inductors or active components.

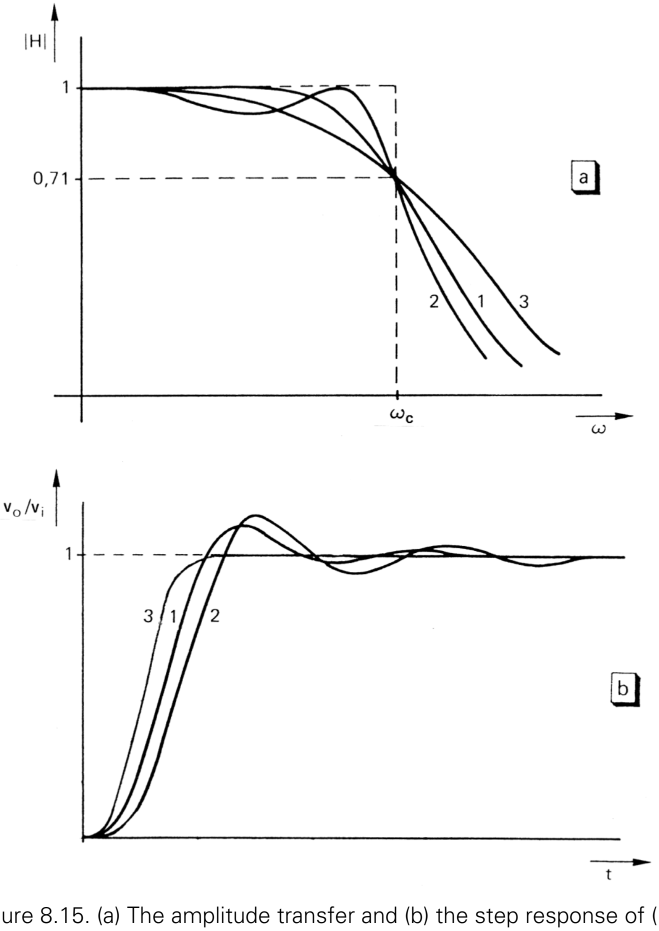


Figure 8.15. (a) The amplitude transfer and (b) the step response of (1) a Butterworth filter, (2) a Chebychev filter and (3) a Bessel filter, all of the 4th order.

So far we have only considered the filter amplitude transfer. However, phase transfer can also be of importance, particularly for filters of a high order. The higher the order, the larger the phase shift will be (Section 6.1.2). There is a relation between the amplitude transfer and the phase transfer of a linear network: it is called the Bode relation. So, having completed a filter design based on specified amplitude transfer, it becomes necessary to check the resulting phase transfer. This may be of paramount importance, in particular in feedback systems, where improper phase characteristics may endanger the system's stability.

SUMMARY

First and second order RC-filters

- There are four basic filter types: lowpass, highpass, bandpass and band reject or notch filters.
- The amplitude transfer of a first-order lowpass RC-filter (integrating network) with time constant τ is approximated by its asymptotes 1 (0 dB) for $\omega\tau << 1$ and $1/\omega\tau$ (–6dB/octave) for $\omega\tau >> 1$. The phase transfer runs from 0 via $-\pi/4$ at $\omega\tau = 1$ to $-\pi/2$.
- The amplitude transfer of a first-order highpass RC-filter (differentiating network) with time constant τ is approximated by its asymptotes $\omega\tau$ (+6dB/octave) for $\omega\tau << 1$ and 1 (0dB) for $\omega\tau >> 1$. The phase runs from $\pi/2$ via $\pi/4$ (at $\omega\tau = 1$) to 0.
- When applying a low-pass or high-pass RC-filter, the influence of the source impedance and the load impedance must be taken into account.
- A band-pass filter, composed of a low-pass and a high-pass RC-filter has an amplitude transfer that has slopes of + and −6dB/octave. The selectivity is poor.
- A double-T or bridged-T filter is an example of an RC-notch filter. Under certain conditions, the transfer is zero for just one frequency.

Filters of higher order

- The amplitude transfer of an n order filter, composed of n cascaded, perfectly matched sections, has a slope of $6n$ dB per octave. The attenuation at the cut-off frequency is $(\frac{1}{2}\sqrt{2})^n$ or $-3n$ dB.
- A (lowpass or highpass) Butterworth filter has a maximally flat transfer in the pass band.
- A Chebychev filter has a maximally steep slope from the cut-off frequency at a given order. Its pass band transfer is oscillatory, the oscillations being more pronounced at higher order.
- A Bessel filter has an optimized step response; its phase transfer is linear up to the cut-off frequency.

EXERCISES

First and second order RC-filters

8.1 The time constant of a passive first-order low-pass filter is $\tau = 1$ ms. Find the modulus and the argument of the transfer for the following frequencies:

a. $\omega = 10^2$ rad/s,
b. $\omega = 10^3$ rad/s,
c. $\omega = 10^4$ rad/s.

8.2 A passive first-order high-pass filter has a time constant $\tau = 0.1$ s. For which frequency is the transfer equal to:

a. 1;
b. 0.1;
c. 0.01?

8.3 A measurement signal with a bandwidth of 0-1 Hz receives interference from a sinusoidal signal of 2 kHz. One therefore tries to attenuate the interference signal by a factor of 100, using a first order filter. The measurement signal itself should not be attenuated by more than 3%. Calculate the limits of the filter's time constant.

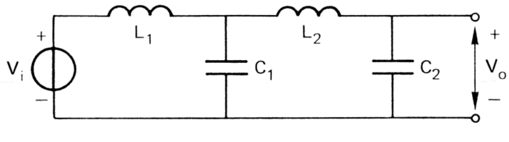
8.4 A periodic, triangular signal contains a third harmonic component whose amplitude is factor 9 below that of the fundamental (even harmonics fail). This signal is applied to a low-pass filter to reshape the signal into a sine wave. Find the resulting sine-wave output distortion (the ratio of the amplitudes of the third harmonic and the fundamental) for filters of orders:

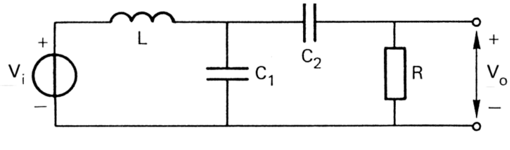
a. 1;
b. 2;
c. 3.

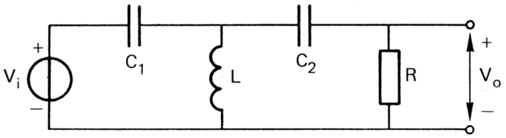
8.5 A measurement signal with frequency 10 Hz, an interference signal with the same amplitude and frequency 50 Hz must be separated: the interference signal must be suppressed by at least a factor of 100 compared to the measurement signal. Find the minimum order to meet this requirement.

Filters of higher order

8.6 Find the type of filter characteristic (low-pass, high-pass, band-pass, notch) for the filters a-c depicted below.







8.7 The modulus of the transfer function of an nth-order low-pass Butterworth filter is

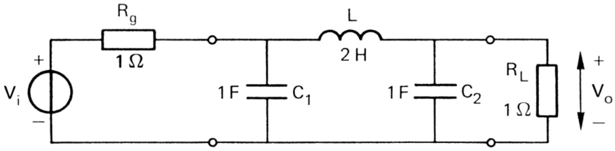
$$H = \frac{1}{\sqrt{1+\left(\dfrac{\omega}{\omega_c}\right)^{2n}}}$$

Calculate the amplitude transfer of this filter for the frequency $\omega = \frac{1}{2}\omega_c$, and express this transfer in terms of dB for

a. $n = 2$;

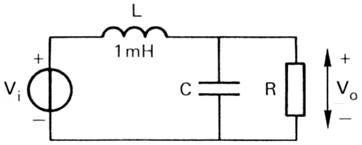
b. $n = 3$.

8.8 Given a third-order LC-filter with source resistance and load resistance.



Calculate the modulus of the complex transfer function, and show that this is a Butterworth filter.

8.9 The inductor in the filter given below has a value $L = 1\text{mH}$.



Find the values of C and R such that the −3dB frequency is 10^5 rad/s, and the characteristic satisfies the Butterworth condition.

9 PN-diodes

Semiconductors are vital to the field of electronics. All active components, with some exceptions, are based on the particular properties of semiconducting materials. It is notably the element silicon that is extensively used because it allows many electronic components to be integrated into a small piece of this material. Other semiconducting materials currently in use are germanium (the first transistor was made from this but nowadays it is used for high frequency signal processing) and gallium arsenide (used for optical components). The first part of this chapter deals with the basic concepts of semiconductors, in particular silicon. Beyond that the operations and characteristics of a particular semiconductor device, the pn-diode, is discussed. The second part illustrates how such pn-diodes are applied to a number of signal processing circuits.

9.1 The properties of pn-diodes

9.1.1 The operation of pn-diodes

Pure silicon (also known as intrinsic silicon) has a rather low conductivity at room temperature, the concentration of free charge carriers (Section 7.1) is small. Once specially technologically treated, the charge carrier concentration can be increased to a predefined level, thus resulting in accurately known conductivity. This process, which is known as doping, involves adding impurities to the silicon crystal.

Silicon is a crystalline material consisting of a rectangular-shaped, symmetric lattice of atoms. A silicon atom has four electrons in its valance band, each of which contributes to the covalent bond with one of the four neighboring atoms. Only very few electrons have enough energy to escape from this bond. When they do, they leave an empty space known as a hole.

In a crystal doped with atoms which has five electrons in its valance band, four of them will be used to form the covalent bonds with four neighboring silicon atoms. The fifth will only be very weakly bound and therefore able to move freely through the lattice. Notice that these electrons do not leave a hole. What they do in any case is contribute to the concentration of free charge carriers in the material in question which is why they are called free electrons. The material that is responsible for this supplying of free carriers is called a donor, the resultant doped silicon is said to be n-type silicon, because the electrons have negative charge. At room temperature almost every impure atom generates an additional free electron. The free charge carrier concentration is therefore equivalent to the donor concentration.

When silicon is doped with atoms that have only three electrons in the valance band, all three will contribute to the bonding with neighboring silicon atoms, there will even be one link missing. This empty place, the hole, can easily be filled by a free electron, if available. The material that facilitates this easy intake of free electrons is called an acceptor. The resulting doped silicon is p-type silicon because of the shortage of electrons (equivalent to holes). There are approximately as many holes as there are impurity atoms.

At first sight, one may think that the conductivity of p-type silicon is even lower than that of intrinsic material. This is not, however, the case at all. When a piece of p-type silicon is connected to a battery, electrons at the negative pole of the battery will drop into the surplus of holes. Due to the electric field produced by the battery they will move from hole to hole and converge on the positive pole. Obviously, this type of conducting differs essentially from the conducting of free electrons. Any time an electron moves to another hole it leaves a new hole, thus resulting in a migration of holes in the opposite direction. To distinguish between conduction in n-type and p-type silicon let us think of holes as positive charge carriers, free to move through the silicon. The mobility of electrons differs somewhat from that of holes. That is why the conductivity of n-type silicon differs from that of p-type, even when equally doped and kept at the same temperature.

What happens when a piece of p-type silicon is connected to a piece of n-type silicon? This pn-junction, which has very particular properties, will play a key role in almost all electronic components. In modern technology, the junction is not made by simply putting two materials together (as was done in the early days of pn-junction formation) but rather by merely partially doping n-type material with acceptor atoms or p-type material with donor atoms. We will now discuss the main properties of such a pn-junction.

Consider the junction of p and n-type silicon at the moment of connection. The p-type contains a high concentration of holes but in the n-type there are hardly any holes. Due to this sharp gradient, holes will diffuse from the p-type to the n-type material. For the same reason, electrons will drift from the n-type to the p-type region. However, when an electron meets a hole, they will recombine and nothing will be left. The result of this recombining is a thin layer on either side of the junction that contains neither free electrons nor free holes. It is depleted of free charge carriers which is why that region is called the depletion region or the depletion layer (Figure 9.1).

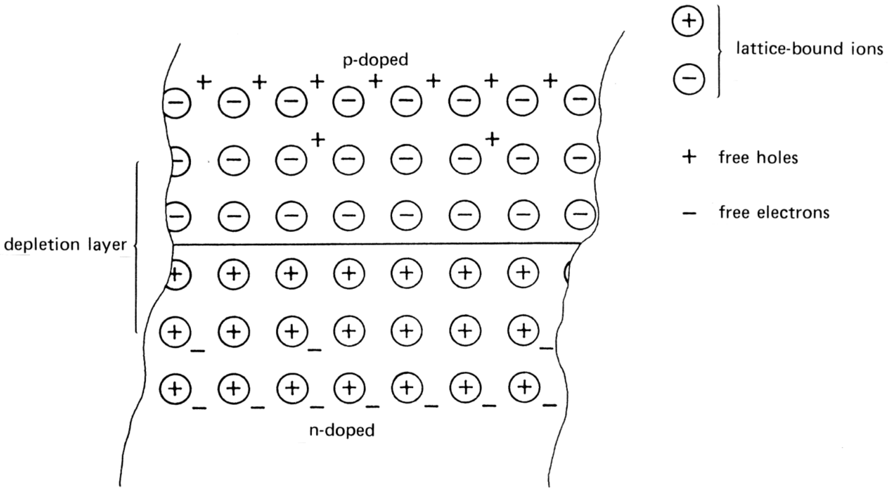


Figure 9.1. On either side of a pn-junction there is a region without free charge carriers known as the depletion region.

The depletion region then ceases to be a neutral zone. When the electrons have drifted away positive ions remain in the n-type material while negative ions remain in the p-type. There is a positive and a negative space charge region on both sides of the junction. Such a dipole space charge is accompanied by an electric field directed from the positive to the negative region or, from the n-type to the p-type side. This electric field prevents the further diffusion of electrons and holes. At a certain depletion layer width an equilibrium will be struck between the electric field strength and the tendency of free charge carriers to reduce any concentration gradient.

The width of the depletion layer depends on the concentration of free charge carriers in the original area. The width can also be influenced by an external electric field. Now suppose that we connect the n-side of the junction to the positive pole of a battery and the p-side to the negative pole. The external field will then be running in the same direction as the internal field across the junction. Moreover, since the depletion region has a very low conductivity (a lack of free charge carriers), the external field will almost completely cover the junction itself, thus reinforcing the original field. We have seen that there is equilibrium concerning the depletion width and the electric field. When an external field is applied (which is added to the internal field) a new equilibrium will develop, this time at a wider depletion layer.

In the case of an external field with opposite polarity the reverse holds: the internal field decreases in accordance with the decrease in the width of the depletion layer.

As the depletion layer is almost devoid of free charge carriers it behaves like an isolator letting no current pass the junction. However, when the external electric field is increased until it fully compensates the internal field, a current can then flow through the material because the conductivity of the doped silicon is quite low.

On the basis of the considerations above it becomes possible to draw the voltage-current characteristic of the pn-junction (Figure 9.2). If we disregard the leftmost part of the characteristic (at V_z), we see that current can only flow in one direction. This element is therefore called a diode, in allusion to the similar property of the vacuum diode.

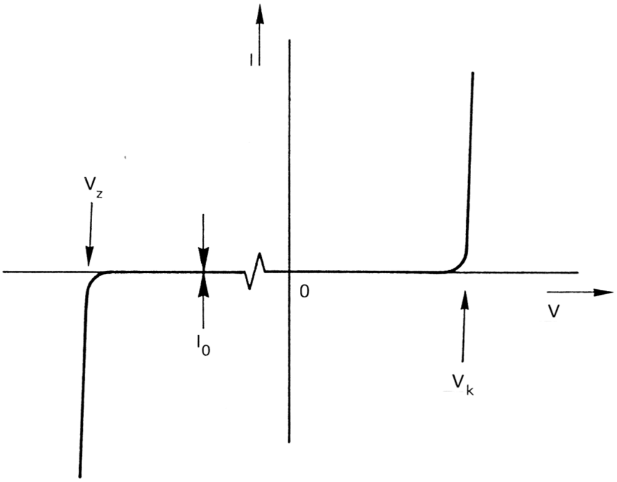


Figure 9.2. The voltage-current characteristic of a pn-diode. V_Z is the Zener voltage, I_0 the leakage current or reverse current and V_k the threshold voltage.

The theoretical relationship between the current through a diode and the voltage across it is given as (9.1):

$$I = I_0\left(e^{\frac{qV}{kT}} - 1\right) \tag{9.1}$$

with q the electron charge (1.6×10^{-19} C), k Boltzmann's constant (1.38×10^{-23} J/K) and T the temperature in K. At room temperature ($T = 300$ K) the factor q/kT equals about 40 V^{-1}. For a negative voltage (reverse voltage), qV/kT is small compared to 1, the current is thus about $-I_0$, and is called the reverse current or diode leakage current. For positive voltages (forward voltage) the current approximates $I \approx I_0 e^{qV/kT}$ which increases exponentially as forward voltage increases by amount e (≈2.78) per 1/40 V = 25 mV. One can roughly assert that due to this strong current rise triggered by the voltage, the diode will conduct above a certain voltage (the threshold voltage V_k in Figure 9.2). Below that voltage the current is almost zero. V_k depends on the material; for silicon it has a value of between 0.5 and 0.8 V. In practice, a value of 0.6 V is used as a rough indication. The reverse current I_0 depends on the material and very much on the temperature. Silicon diodes intended for general use have a leakage current below 10^{-10} A which increases exponentially as the temperature increases by a factor of 2 per 6 to 7 °C. Conversely, at constant current, the voltage across the diode decreases by about 2.5 mV as the temperature increases 1 °C.

From Figure 9.2 and expression (9.1) it appears that a pn-diode is a non-linear element: it is not possible to describe a diode in terms of a single resistance value. To characterize the diode in more simple terms than the exponential relation of (9.1), we will introduce the differential resistance, defined as $r_d = \mathrm{d}V/\mathrm{d}I$. From (9.1) it follows that $V = (kT/q)\ln(I/I_0)$, the differential resistance of the diode is therefore kT/qI, which is inversely proportional to the current I. At room temperature $kT/q = 1/40$ and so the differential resistance of the diode is about 25 Ω for a current $I = 1$ mA. This value is independent of the construction, and it is the same for all diodes. Using this rule of thumb it becomes easy to find, for an arbitrary current, the differential resistance.

The reciprocal value of the differential resistance is the (differential) conductance: $g_d = 1/r_d$, which corresponds to the slope of the tangent at a point of the characteristic given in Figure 9.2. The conductance at 1 mA is 40 mA/V which is directly proportional to the current.

Other important characteristics of a diode are the maximum possible current (in a forward direction) and voltage (in a reverse direction). The maximum current is about 10 mA for very small types and may be several kA for power diodes. The maximum reverse voltage ranges from a few V up to several 10 kV. At maximum reverse voltage, the reverse current increases sharply (see the left part of Figure 9.2), due to a breakdown mechanism. One such mechanism is Zener breakdown, an effect that is employed in special diodes, called Zener diodes that are able to withstand breakdown. In the breakdown region these diodes have a very low differential resistance (the slope of the I-V characteristic is very steep): the voltage hardly changes at varying current. This property is used for voltage stabilization. For applications that require a very stable Zener voltage, diodes are constructed with an additional diode for temperature compensation; the net temperature coefficient of the Zener voltage can be as low as 10^{-5} per °C. Zener diodes are constructed for voltages ranging from about 6 V to several 100 V. Figure 9.3 presents the symbols of a normal diode and a Zener diode.

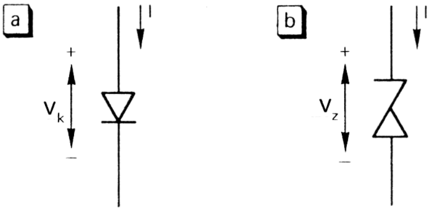


Figure 9.3. Symbols for (a) a diode, (b) a Zener diode.

9.1.2 Photodiodes

The leakage current (reverse current) of a diode originates from the thermal generation of free charge carriers (electron-hole pairs). Charge carriers that are produced within the depletion layer of the diode will drift away because of the electric field: electrons drift to the n-side and holes to the p-side of the junction. Both currents contribute to an external leakage current, I_0. The generation rate of such electron-hole pairs depends on the energy of the charge carriers: the more energy there is, the greater the number of electron-hole pairs produced will be. It is possible to generate extra electron-hole pairs by adding optical energy. If light is allowed to fall on the junction, free charge carriers will be created that will increase the leakage current of the diode. Diodes that are

designed to employ this effect (light sensitive diodes or photodiodes), have a reverse current that is almost proportional to the intensity of the incident light.
The main characteristics of a photodiode are:

- spectral response, expressed as ampère per watt or ampère per lumen (Figure 9.4a). Silicon photodiodes have a maximal response for light with a wavelength of about 800 nm,

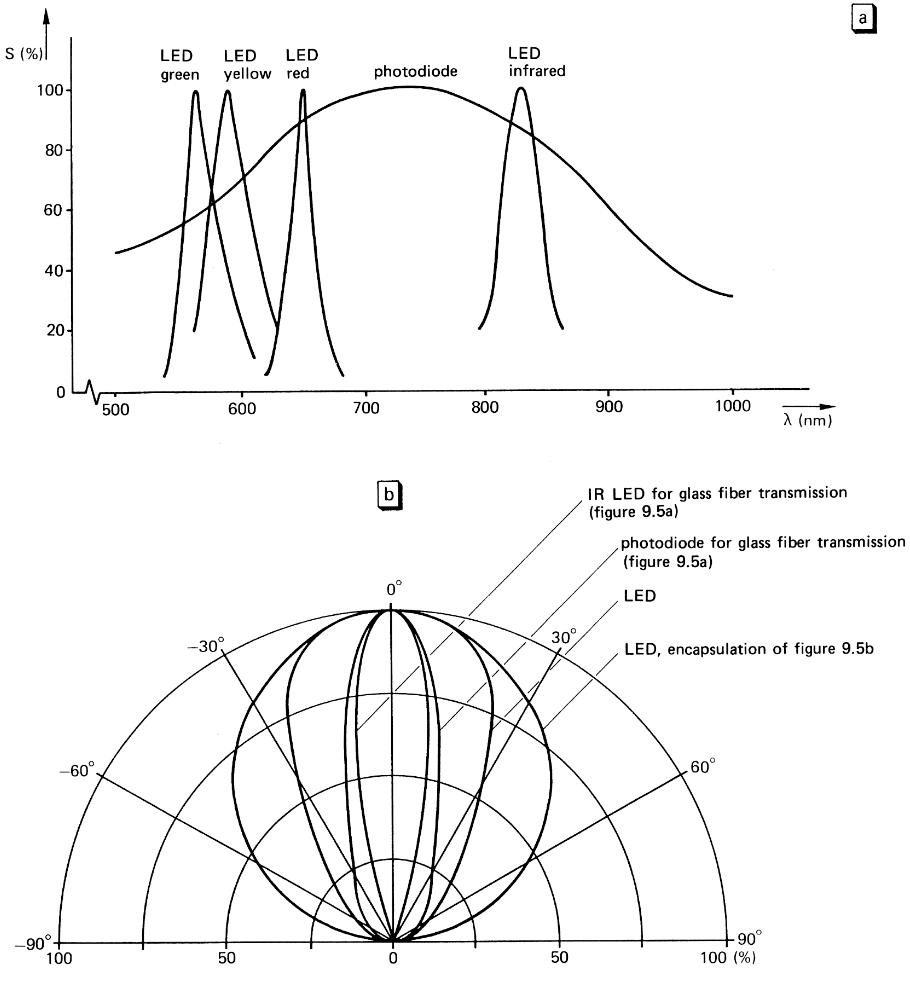


Figure 9.4. (a) Spectral response S and (b) a polar sensitivity diagram of a photodiode and several light-emitting diodes. The shape of the polar diagram depends on the device's encapsulation (i.e. on whether it does or does not have a built-in lens).

- dark current, the reverse current in the absence of light. As can be expected from the nature of the reverse current, the dark current of a photodiode increases markedly as temperatures rise. Usually this increase is great compared to that of the reverse current of normal diodes and ranges from several nA to μA, depending on the surface area of the device,

- quantum efficiency, that is to say, the ratio between the number of optically generated electron-hole pairs and the number of incident photons. This efficiency is better than 90% at the peak wavelength.

Photodiodes have a much faster response to light variations than photoresistors (Section 7.2.1.2), they can follow light signals with a bandwidth of several 100 MHz, and are, therefore, suitable for glass fiber transmission applications (Figure 9.5a)

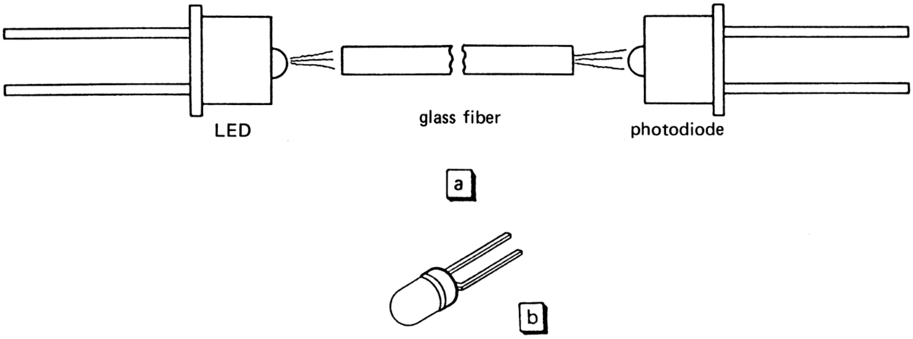


Figure 9.5. (a) An LED and a photodiode applied to glass fiber communication links, (b) common type LED.

9.1.3 Light-emitting diodes (LEDs)

Some semiconductor materials generate photons when the forward current through a pn-junction exceeds a certain value. Such a diode is called a light-emitting diode or LED. Silicon is not a suitable material for this effect. A semiconductor that is better in this respect is gallium arsenide, possibly supplemented with small amounts of phosphorus, aluminum or indium. The color of an LED is determined by the composition of the semiconductor. The spectrum of available LEDs ranges from infrared to blue.

The main characteristics of an LED are:

- its peak wavelength which, depending on the type of material used, is between 500 nm (blue) and 950 nm (infrared);
- the polar emissivity diagram. There are LEDs that produce very narrow beams (usually by applying a built-in lens) or wider beams (Figure 9.4b). Narrow-beam LEDs are suitable for coupling with glass fibers (Figure 9.5a);
- the maximum allowable current, which also determines the maximal intensity. The peak current is around 100 mA and the corresponding forward voltage is 1 to 2 V.

LEDs are constructed in various encapsulations, with or without a lens, with only one element or a whole array. They are widely used as alpha-numeric displays. Figure 9.5b shows a simple type. There are also LEDs that emit two colors, depending on the direction of the current. This device consists of two independent LEDs in a single encapsulation connected in an anti-parallel fashion.

9.2 Circuits with pn-diodes

PN-diodes can be used in one of two ways for non-linear signal processing. The first possibility is based on the exponential relationship between the voltage and the current. This creates circuits for exponential and logarithmic signal converters and for analog multipliers. Such circuits will be discussed in Chapter 13. Looking to the diode characteristic of Figure 9.2, it appears that for voltages below V_k the diode behaves like a very large resistance. When conducting, the (differential) resistance is low (25 Ω around the 1 mA point). Furthermore, the voltage across the diode remains almost V_k irrespective of the forward current. Similarly, with negative currents, the voltage equals the Zener voltage V_Z. So, a diode behaves like an electronic switch: when conducting, the switch is closed, and acts as a short with a series voltage equal to V_k in the forward mode and V_Z in the reverse mode (only for Zener diodes). If the voltage is below V_k, the switch will be off and the current will be zero. This rough approximation is used in this particular section to analyze a number of commonly used pn-diode circuits.

9.2.1 Limiters

A limiter or clipper is a circuit that has a prescribed limited output voltage. Figure 9.6a shows a circuit for the limitation up to a maximum voltage and Figure 9.6b gives the corresponding transfer characteristic.

As long as the input voltage $v_i < V_k$, the current through the diode will be zero so the output voltage will be $v_o = v_i$. When v_i reaches V_k, the diode becomes forward biased and conducts but its voltage remains V_k. The diode current equals $(v_i - V_k)/R$, so with R the input current can be limited as well.

When the diode connections are reversed, the output voltage is limited to a minimum value equal to $-V_k$. By connecting two diodes in an anti-parallel fashion (Figure 9.6c), v_o is limited to between $-V_k$ and $+V_k$ (Figure 9.6d). For overload protection such limiting circuits are often connected across the input of a sensitive measurement system.

Limiting up to a voltage other than V_k is achieved using Zener diodes (Figure 9.7).

The maximum output voltage is the voltage for which both diodes conduct. The voltage across Z_1 is the Zener voltage V_{Z1} and that across Z_2 is the normal voltage drop V_k. A minimum voltage develops when both diodes are conducting in the reverse direction: the voltage across Z_1 is the forward biased voltage V_k and the voltage over Z_2 is its Zener voltage V_{Z2}. The output limits are not affected by load resistance (Figure 9.7b). Between the two output limits (when both diodes are reverse biased), the transfer of the circuit is $v_o = v_i R_L/(R + R_L)$ which is the formula for the normal voltage divider circuit.

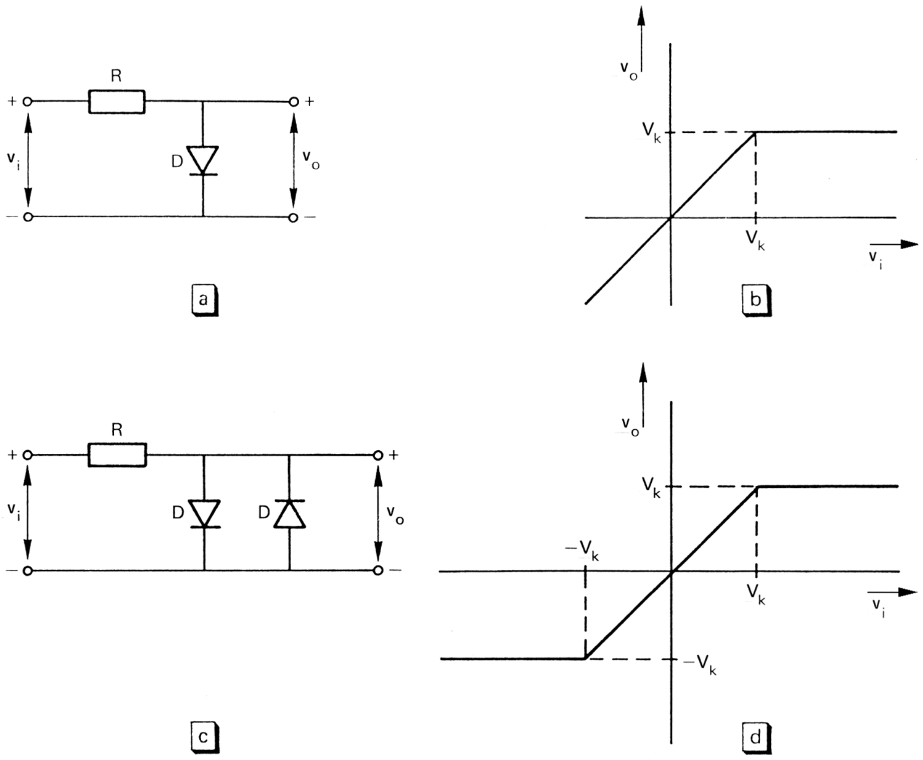

R
+
+
v_i
D
v_o
−
−
a
v_o
V_k
V_k
v_i
b
R
+
+
v_i
D
D
v_o
−
−
c
v_o
V_k
$-V_k$
V_k
v_i
$-V_k$
d

Figure 9.6. (a) A limiter for limitation to a maximum voltage, (b) the corresponding transfer characteristic, (c) a limiter for both positive and negative maximum voltages, (d) the corresponding transfer characteristic.

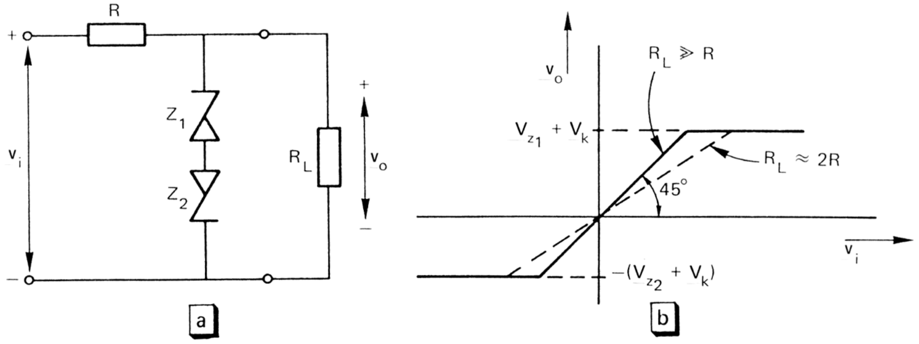

R
+
Z_1
Z_2
v_i
R_L
+
v_o
−
−
a
v_o
$R_L \gg R$
$V_{z_1} + V_k$
$R_L \approx 2R$
45°
v_i
$-(V_{z_2} + V_k)$
b

Figure 9.7. (a) A limiter with Zener diodes, (b) the transfer characteristic for two values of the load resistance R_L; $V_{Z1}<V_{Z2}$.

9.2.2 Peak detectors

A peak detector is a circuit that creates a DC output voltage equal to the peak value of a periodic input voltage. Figure 9.8a shows a peak detector circuit in its most simple form with a capacitor and a diode.

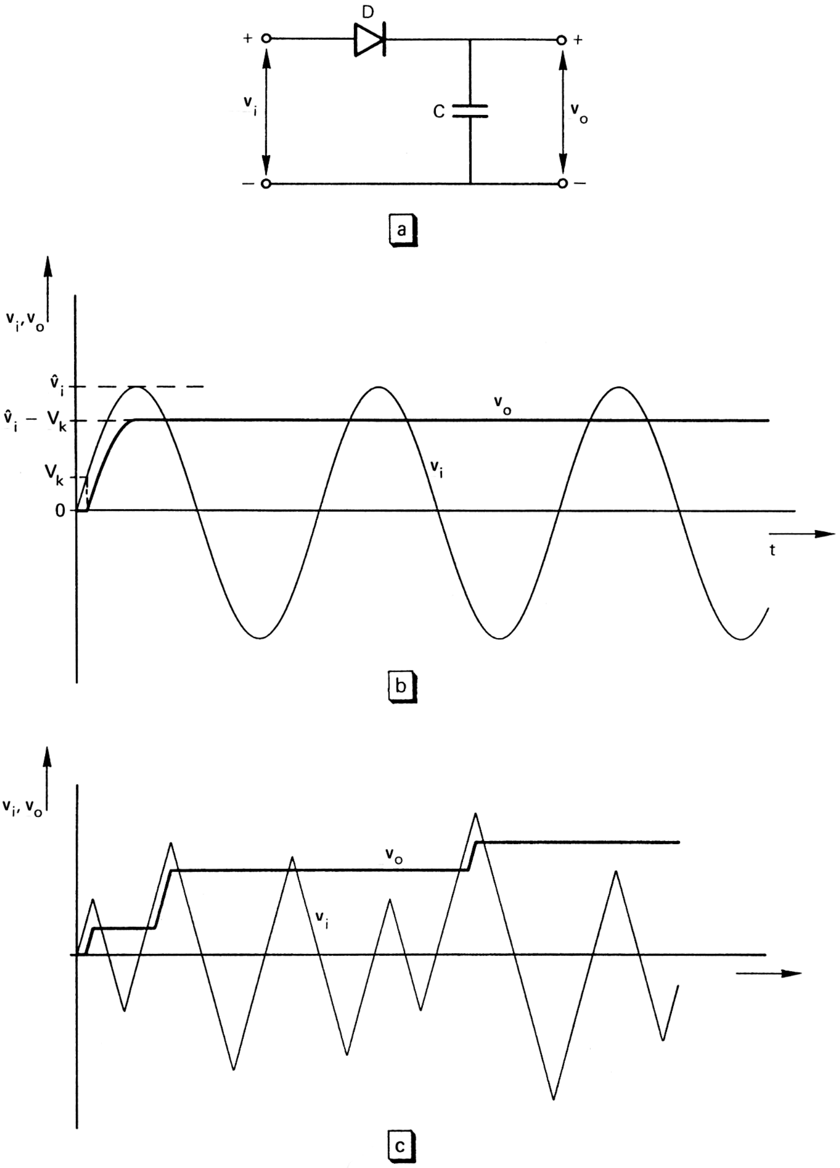


Figure 9.8. (a) A diode peak detector for positive peak values, (b) output voltage v_o for sinusoidal input voltage v_i, (c) output voltage v_o for triangular input voltage v_i.

To explain how this circuit operates let us start with an uncharged capacitor and consider the rising part of the input voltage (Figure 9.8b). At $t = 0$, the voltage across the diode is zero: the diode is reverse biased. The output v_o is also zero and keeps to that value until the input voltage v_i reaches V_k. Only then does the diode reach a state of conduction so that a current can flow and the voltage can remain V_k. The output thus keeps up with the input and there is a constant difference of V_k. The capacitor is charged by a positive current through the diode. The current cannot flow into the inverse direction and so the capacitor cannot be discharged. When the input voltage decreases after having reached its peak value the output voltage will therefore remain constant. Its value equals the peak value of the input minus V_k. This situation will be maintained as long as the input voltage does not again exceed $v_o + V_k$. Only when the value is slightly higher at the next peak will the diode conduct. The capacitor, though, will be charged up to the new peak value in the way depicted in Figure 9.8c for triangular input voltage.

The peak detector seen in Figure 9.8a can only detect the absolute maximum ever. If we want the circuit to also respond to the peak value of a gradually decreasing amplitude then the capacitor must partly discharge between two successive peaks. This is achieved by connecting a resistor across the diode or the capacitor (Figure 9.9a). Even when the diode is reversed biased the capacitance will discharge, resulting in a small decrease in the output voltage (Figure 9.9b). The capacitor is charged again at each new maximum that exceeds $v_o + V_k$. In other words, also when amplitude is gradually decreasing, the circuit can follow the peak value of a periodic signal. The price of this simple measure is a small ripple in the output signal, even at constant amplitude. To estimate an appropriate resistance value for a minimum output ripple, we must assume that there is a linear discharge curve (instead of a negative exponential curve). At $t = 0, T, 2T, ...,$ v_o equals $\hat{v}_i - V_k$ and falls with a rate $\hat{v}_i/\tau$ (V/s). The output ripple is:

$$\Delta v = \left(\hat{v}_i - V_k\right) - \left[\left(\hat{v}_i - V_k\right)\frac{\tau - T}{\tau}\right] = \left(\hat{v}_i - V_k\right)\frac{T}{\tau} \tag{9.2}$$

The true value is somewhat smaller than this approximation, because v_o starts to rise again from the moment slightly before $t = T, 2T, \ldots$ To minimize the ripple, τ must be large. To keep track of a decreasing amplitude, τ must be small. A compromise must therefore be made between ripple amplitude and the response time to decreasing amplitudes.

The output voltage of a peak detector is always about V_k below the actual peak value. When changing the polarity of the diode in Figure 9.8a, the output voltage responds to the negative of the peak value: $v_o = \hat{v}_i + V_k$. An AC voltage with an amplitude that is below V_k (about 0.6 V) cannot be detected with this simple circuit.

9.2.3 Clamp circuits

A clamping circuit moves the average level of an AC signal up or down so that the top has a fixed value. Figure 9.10 shows the most simple circuit configuration and gives an example of sinusoidal input and output voltage.

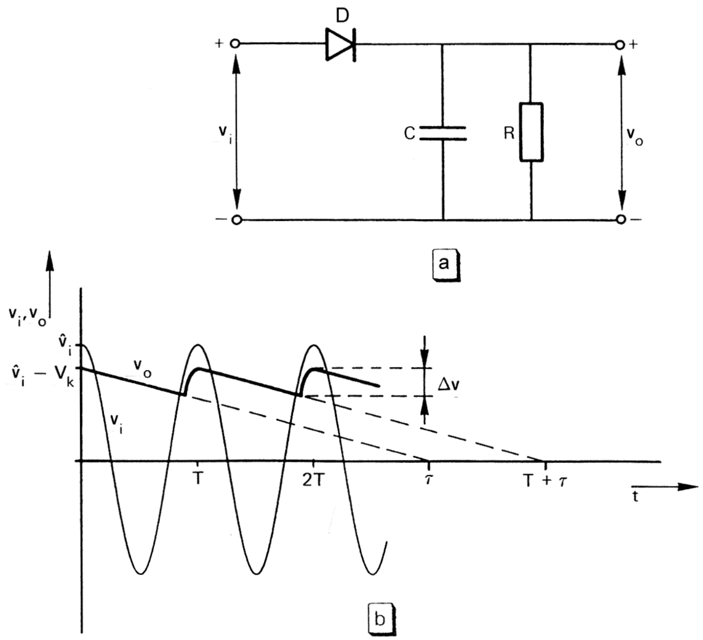


Figure 9.9. (a) Peak detector with discharge resistance in order to allow a slowly decreasing amplitude to be detected, (b) the determining of the ripple voltage.

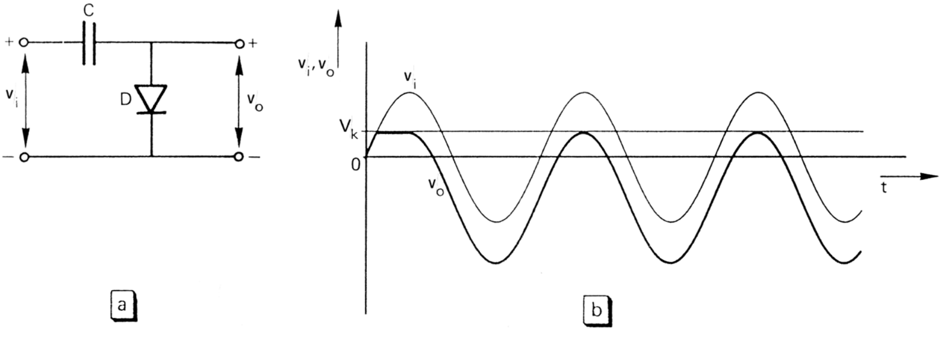


Figure 9.10. (a) A clamping circuit, (b) the response to a sinusoidal input signal.

At $t = 0$ it is assumed that the capacitor is uncharged. Obviously, the output voltage can never exceed value V_k. In addition, the capacitor can never be discharged at a diode voltage that is below V_k. With these conditions in mind it becomes clear precisely how this circuit operates. At increasing input v_i, the diode remains reverse biased, no current flows, so $v_o = v_i$. As soon as v_i reaches V_k the diode will conduct and behave like a short-circuit: $v_o = V_k$. This situation is maintained until the input has reached its maximum. In the meantime, the capacitor is charged up to a voltage of $v_c = \hat{v}_i - V_k$, but cannot discharge. As a result, the output follows the input and there is a constant difference of just V_k , as long as the output does not exceed the value V_k .

As the capacitor can be charged but not discharged, the circuit only functions well at constant or increasing signal amplitude. This drawback is eliminated by connecting a resistor across the diode or capacitor to allow some discharge of the capacitor between two successive input signal maximums (c.f. the peak detector of Section 9.2.2).

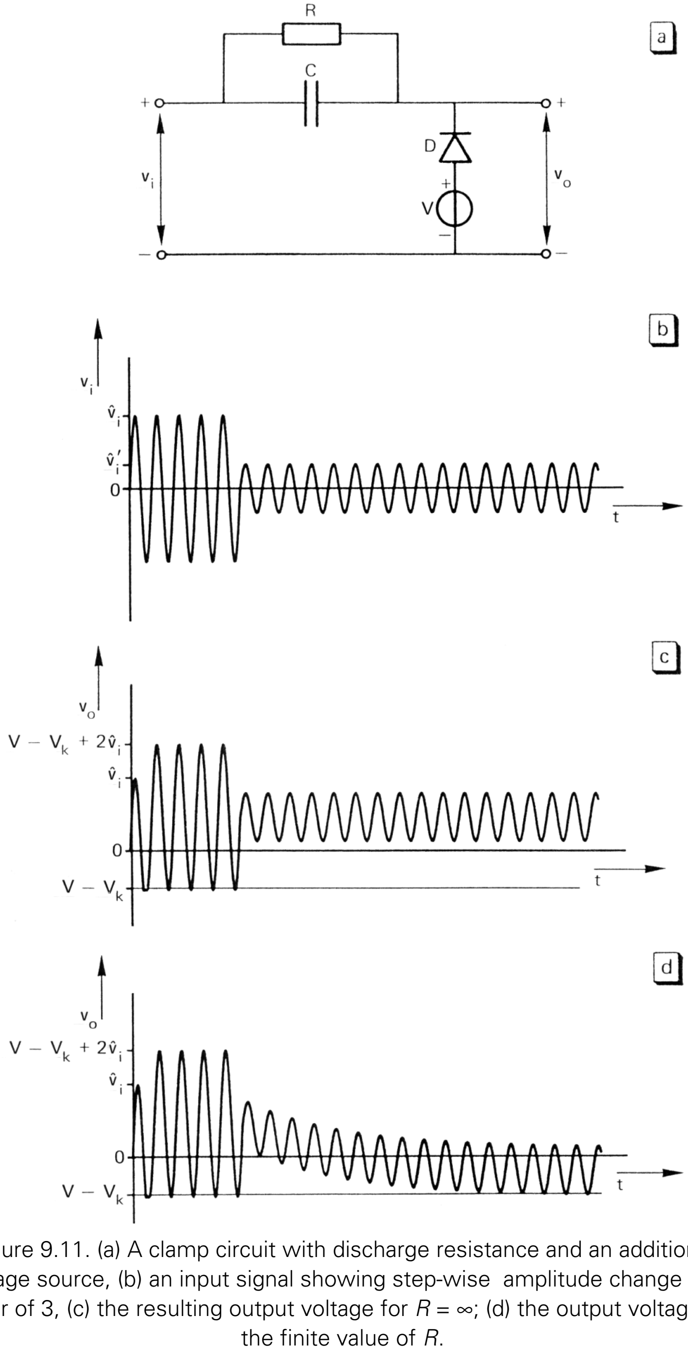


Figure 9.11. (a) A clamp circuit with discharge resistance and an additional voltage source, (b) an input signal showing step-wise amplitude change by a factor of 3, (c) the resulting output voltage for $R = \infty$; (d) the output voltage for the finite value of R.

An arbitrary clamping level can be obtained by using a voltage source in series with the diode. The clamping level becomes $V + V_k$. If the polarity of the diode is changed this will result in clamping on the negative peaks of the input. Figure 9.11 shows a circuit that clamps the negative peaks at a level of $V + V_k$ and which also responds to slowly decreasing amplitudes as depicted in Figure 9.11b. Clamping circuits are used to set a proper DC level for AC signals. This can be combined with a peak detector (Figure 9.8a) to create a peak-to-peak detector where the output equals the peak-peak value of the input, minus $2V_k$.

9.2.4 DC voltages sources

Most electronic systems require a more or less stable power supply voltage and sometimes a very stable and precisely known reference voltage. The voltage should not be affected by the current through it and made to behave like an ideal voltage source. Special integrated circuits are available for high stability requirements which give a stable output voltage for various output currents over a wide power range. Simpler methods can be adopted in situations where the requirements are less strict. In this section we shall introduce two DC voltages sources, the first uses an unstabilized DC voltage as a primary power source while the second uses an AC power source.

A voltage stabilizer with Zener diodes

As explained in Section 9.1, the Zener voltage is virtually independent of the reverse current through the diode. This property is employed when realizing a simple yet fairly stable voltage source (Figure 9.12).

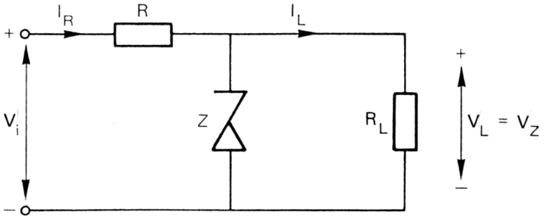


Figure 9.12. Voltage stabilization using a Zener diode.

The input of this circuit is a voltage V_i that may vary over a substantially wide range. The diode is biased by its reverse current, the value of which is laid down by resistor R. Obviously, under the condition $V_i > V_Z$, the output voltage V_L equals V_Z , which makes it almost independent of the current through the diode. When the circuit is loaded with R_L, a current flows through the load equal to $I_L = V_Z / R_L$. As long as $I_R > I_L$, enough current will remain for the Zener diode to maintain its Zener voltage. The output is constant, irrespective of the load resistance. For this particular application there are Zener diodes available that have a maximum tolerated power of 0.3 up to 50 W.

Graetz bridge

Graetz' bridge circuits (Figure 9.13) are used extensively in systems that require a DC supply voltage but where only AC power is available.

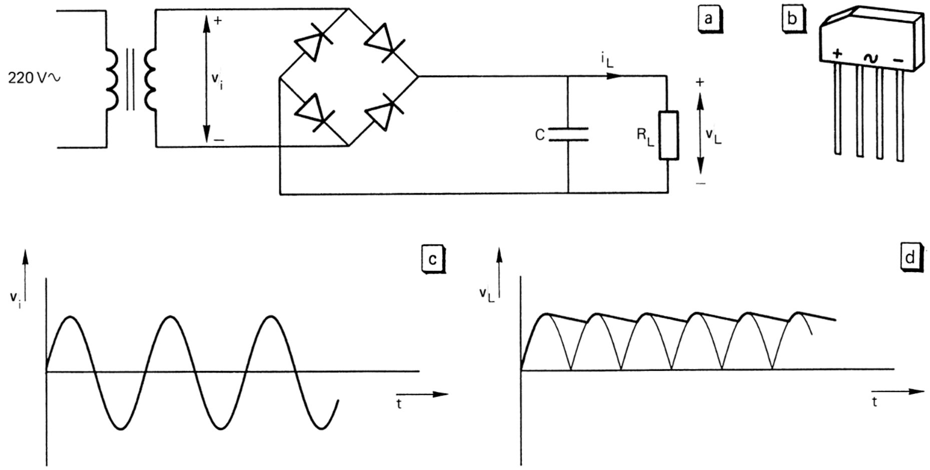


Figure 9.13. (a) A simple power supply circuit consisting of a transformer, a diode bridge and a capacitor, (b) a rectifier bridge in a single housing, (c) the input voltage of the bridge circuit, (d) rectified output where the ripple depends on the load current and the capacitance *C*.

First of all, the primary AC voltage is reduced to a proper value of v_i by the transformer. Due to the four diodes, the current I_L through R_L can only flow in the direction indicated in Figure 9.13a. The polarity of the output voltage v_L is always positive, irrespective of the polarity of v_i (Figure 9.13d, thin line). In conjunction with this property the diode bridge is also sometimes known as a double-sided rectifier (a single-sided rectifier simply clips the negative halves of the sine wave). The capacitor *C* forms a peak detector (Section 9.2.2) together with the bridge's diodes. The ripple in the output signal v_L (Figure 9.13d, bold line) is caused by capacitor discharge through the load resistance. A large *C* value is required to keep this ripple below a specified value, even at high load currents.

Rectifying bridges are available as single components (Figure 9.13b) for a wide range of the allowed maximum current and power.

SUMMARY

The properties of pn-diodes

- The conductivity of n-doped silicon is mainly determined by the concentration of free electrons. In p-doped silicon the conductivity is mainly determined by the concentration of holes.
- The relation between the current through a pn-junction and the voltage is given as $I = I_0(e^{qV/kT} - 1)$. For $V < 0$, $I = I_0$, the leakage current or the reverse current of the diode. This current is very temperature-dependent.

- A silicon diode becomes conductive at a forward voltage of about 0.6 V. At constant current, the temperature sensitivity of the forward voltage is roughly –2.5 mV/K.
- The differential resistance of a pn-diode is $r_d = kT/qI$; r_d is about 25 Ω at $I = 1$ mA.
- Zener diodes show an abrupt increase in the reverse current at the Zener voltage. They are used as voltage stabilizers and voltage reference sources.
- The leakage current of light-sensitive diodes or photodiodes is proportional to the intensity of the incident light. To operate properly the diode should be reverse biased.
- Light-emitting diodes or LEDs emit a light beam, the intensity of which is roughly proportional to the forward current. The color or wavelength of the light is determined by the composition of the semiconductor material.

Circuits with pn-diodes

- A pn-diode can be used as an electronic switch: for $V_d < V_k$, its resistance is very high. When conducting, the resistance is low, and the voltage across the diode is about 0.6 V.
- Diode limiters make use of the property that the diode voltage is limited to about $V_k = 0.6$ V or (with Zener diodes) to the Zener voltage V_Z.
- The peak value of a periodical signal can be measured using the diode-capacitor circuit given in Figure 9.8. The output voltage is 0.6 V below the peak voltage. The response to decreasing amplitudes is improved by using an additional resistor that introduces a ripple voltage at the output.
- With the diode-capacitor circuit shown in Figure 9.10, the top of a periodic signal is shifted to a fixed value, irrespective of the signal amplitude. The response to slowly increasing amplitude is improved by additional resistance.
- With a Zener diode a reasonably stable voltage source can be realized.
- The Graetz bridge is a double-sided rectifier. Using a relatively large capacitor this rectified voltage can be converted into a DC voltage with a small ripple.

EXERCISES

The properties of pn-diodes

9.1 What is the theoretical relationship between the current through a pn-diode and the voltage across it?

9.2 Give the approximate value of the differential resistance of a pn-diode at 1 mA, at 0.5 mA and at 1 μA. Give also the conductance values.

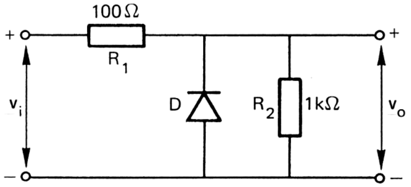
9.3 What is the change in the diode voltage at constant current when there is a temperature increase of 10 °C?

9.4 What is the change in the diode voltage at constant temperature when the current increases by a factor of 10?

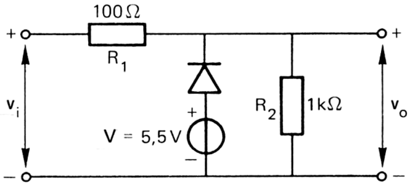
9.5 To obtain the value of the series resistance r_s of a diode the voltage is measured in two different currents: 0.1 mA and 10 mA. The respective results are 600 mV and 735 mV. Find r_s.

Circuits with pn-diodes

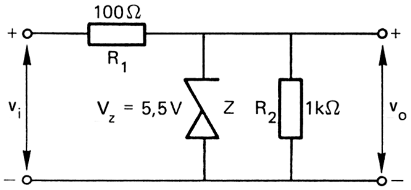
9.6 Draw the transfer function (output voltage versus input voltage) of each of the a-d circuits depicted below. The diode has an ideal *V-I* characteristic; $V_k = 0.5$ V.


100Ω
R1
D
R2
1kΩ
vi
vo
+
−

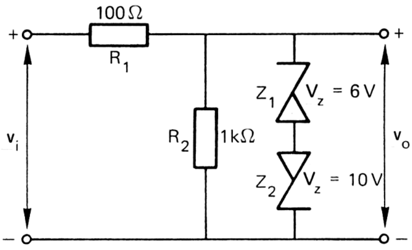
a


100Ω
R1
V = 5,5V
R2
1kΩ
vi
vo
+
−

b

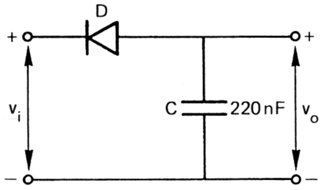

100Ω
R1
Vz = 5,5V
Z
R2
1kΩ
vi
vo
+
−

c


100Ω
R1
R2
1kΩ
Z1
Vz = 6V
Z2
Vz = 10V
vi
vo
+
−

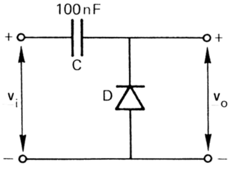
d

9.7 Find the output voltage of the peak detector given below for sinusoidal input voltages with amplitude 6 V, 1.5 V and 0.4 V and zero average value. The diode has ideal properties, $V_k = 0.6$ V.

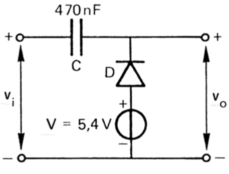

D
C
220nF
vi
vo
+
−

9.8 Refer to the circuit in the preceding exercise *C* =0.1 µF. The input voltage is a sine wave with a frequency of 5 kHz. A resistor *R* is connected in parallel to *C*. Find the minimum value of *R* when the ripple (peak-to-peak value) is less than 1% of the amplitude of v_i. The diode voltage V_k can be neglected with respect to the input amplitude.

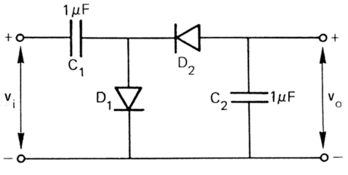
9.9 Find the average value of v_o for each of the following a-c circuits at an input amplitude of 5 V. Assume that $V_k = 0.6$ V.


100nF
C
D
vi
vo
+
−

a

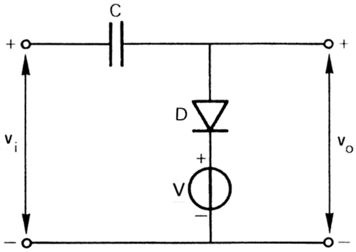

470nF
C
D
V = 5,4V
vi
vo
+
−

b


1µF
C1
D1
D2
C2
1µF
vi
vo
+
−

c

9.10 The figure given below shows a clamp circuit for clamping the negative tops of a signal to an adjustable voltage V. What is the range of V if this circuit is to be properly operated. Bear in mind that the input signal is, on average, a sinusoidal voltage of zero with a peak value value $\hat{v}_i$; $V_k = 0.6$ V.



9.11 Sketch the output signal of a Graetz bridge loaded with a resistor of 20 Ω, and connected to:

a. an AC voltage of 50 Hz, rms value 10 V;

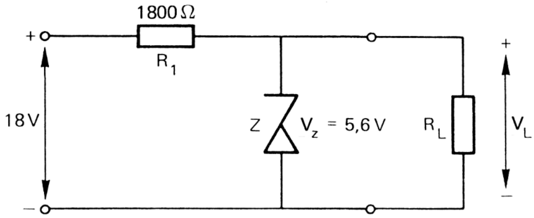
b. a DC voltage of +10 V;

c. a DC voltage of −10 V.

What is the ripple voltage in all three cases? $V_k = 0.6$ V.

9.12 Answer the same questions as those posed in 9.11 but imagine now that a capacitance of 1500 μF is connected to the bridge.

9.13 Find the minimum value of R_L in the following circuit for which the output voltage is just 5.6 V.



10 Bipolar transistors

Bipolar transistors consist of three alternate layers of p-type and n-type silicon. By making the middle layer of the three very thin it becomes possible to amplify electronic signals with this component. In this chapter we shall start by examining the transistor's operating principles before going on to describe its behavior as an active electronic component. Examples will be given of amplifier circuits with bipolar transistors.

10.1 The properties of bipolar transistors

10.1.1 Construction and characteristics

Figure 10.1 presents the schematic structure of a bipolar transistor. In accordance with the various materials used, two types of transistors exist: npn-transistors and pnp-transistors. The same figure also mentions the names of the respective parts (the corresponding terminals have identical names) and shows the circuit symbol for both types.

We shall first consider the transistor as a series of two pn-junctions (or pn-diodes) with one common part. One of the diodes is connected to a reverse voltage, the other to a forward voltage. The common part is called the transistor base, the adjacent part of the reverse biased diode is the collector and the adjacent part of the forward biased diode is the emitter. As the base-emitter diode has forward voltage, a current will flow through that junction. In an npn-transistor this current is composed of electrons flowing from the emitter to the base region. Under normal conditions these electrons leave the transistor via the base connection. No current will in fact flow through the base-collector junction because of the reverse voltage.

The situation changes significantly when the width of the base region is made very thin, as indicated in Figure 10.1. The vast majority of the electrons that enter the base will survive the journey through the base and enter the collector region. Once there, they will be drawn into the collector by the electric field across the base-collector junction. Only a fraction of the electrons fail to reach the collector and leave the transistor via the base terminal. In an npn-transistor where the voltages are properly connected, the electrons flow from the emitter to the collector, thus corresponding to a physical current between the collector and the emitter and thereby passing the forward biased base-emitter junction and the reverse biased base-collector junction.

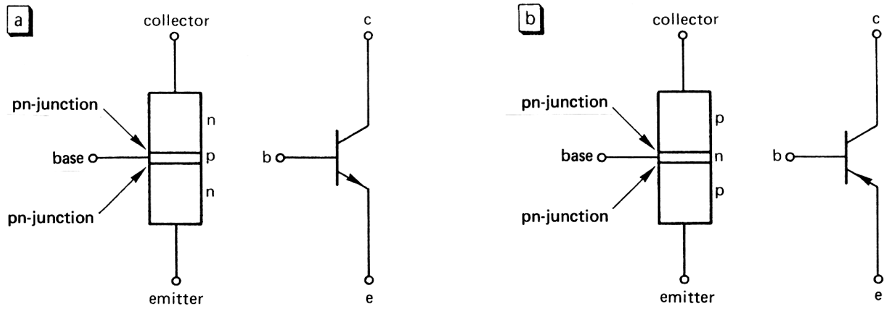


Figure 10.1. (a) Schematic structure and npn-transistor symbol (b) the same for a pnp-transistor.

A pnp-transistor operates in exactly the same way There holes flow from the emitter via the thin base region towards the collector. The polarity of the physical current is positive from emitter to collector because the holes are positive charge carriers.

We have seen that only a fraction of the total current flows through the base. This fraction is an important parameter for the transistor, it determines the transistor's current gain β which is defined as the ratio between the collector current I_C and the base current I_B, so: $I_C = I_B$. The current gain depends to a great extent on the width of the base region and ranges from 100 to 300 for low-power transistors up to 1000 for special types (super-β transistors). High-power transistors have a much lower β, of 20 or even less.

The current through the base-emitter diode (the emitter current I_E) still satisfies the diode equation 9.1. As the collector current is almost equal to the emitter current (the difference is just the small base current) the collector current satisfies (9.1) as well:

$$I_C \approx I_0 e^{qV_{BE}/kT}$$

The collector current is therefore determined by the base-emitter voltage V_{BE}, it is independent of the base-collector voltage V_{BC} under the condition of a reverse biased base-collector junction. In other words, the collector behaves like a current source: the collector current does not depend on the collector voltage. More specifically, it is a voltage-controlled current source, because the current is directly related to the base-emitter voltage. It is also a current amplifier with a current gain of β, because the collector current satisfies the relation $I_C = \beta I_B$.

It is important to distinguish between the two properties of a bipolar transistor: its output signal (the collector current) can be controlled either by an input voltage (V_{BE}) or by an input current (I_B).

Figure 10.2 depicts the typical characteristics of a bipolar transistor: in (a) the relation is shown between the collector current and the base-emitter voltage (at a fixed base-collector voltage) and in (b) the collector current versus the collector voltage (relative to the emitter voltage) is shown with V_{BE} as a parameter.

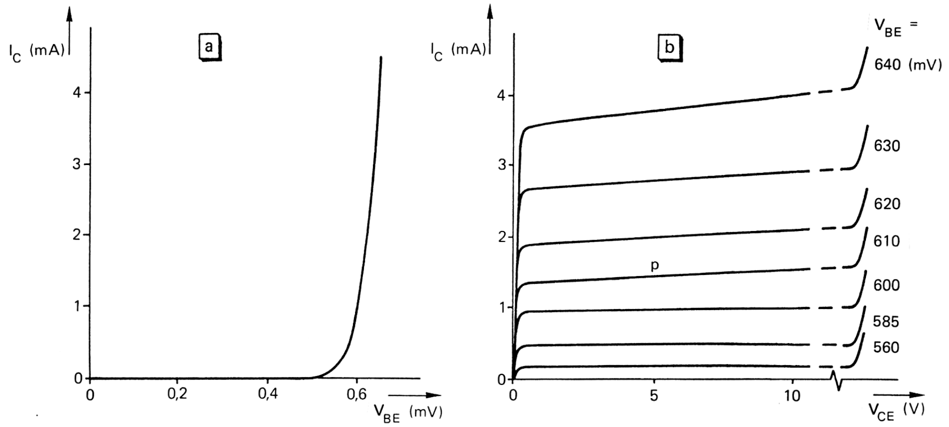


Figure 10.2. The typical characteristics of a bipolar transistor, (a) I_C versus V_{BE} at constant V_{BC}, (b) I_C versus V_{CE} for various values of V_{BE}. Collector breakdown occurs when the reverse voltage of the base-collector junction is increased too much.

Obviously the actual characteristic differs in several respects from the behavior as described previously. Firstly, the collector current (Figure 10.2a) increases less than the exponential relation of the pn-diode would indicate because of the emitter's resistance and base materials. The collector current (Figure 10.2b) is not totally independent of the collector voltage, due to the so-called Early-effect. What is not apparent from Figure 10.2 is that the current gain varies somewhat depending on the collector current: β decreases at very low and very high currents. Finally, a leakage current flows through the reverse biased base-collector diode.

Other important parameters of a bipolar transistor are the maximum reverse voltage for the collector junction, the maximum forward voltage for the emitter junction and the maximum power $I_C \cdot V_{CE}$. The maximum power ranges from roughly 300 mW for the smaller types to over 100 W for power transistors with forced cooling. Figure 10.3 shows different transistors in various encapsulations.

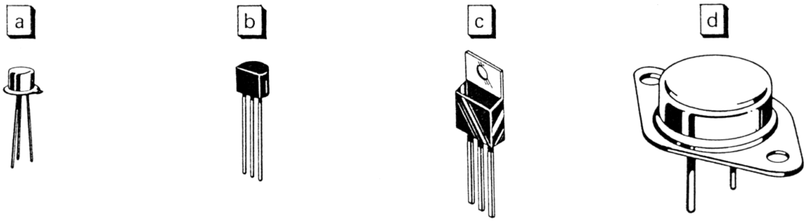


Figure 10.3. Some examples of various transistors, (a) & (b) low-power transistors (up to about 400 mW), (c) medium power (about 10 W), (d) high power transistor (about 100 W).

10.1.2 Signal amplification

If a bipolar transistor is to be employed for linear signal amplification it has to be correctly biased. This means that all voltages between the terminals and consequently all the currents through the terminals should have a proper value. This state is called the bias point or operating point of the transistor, look for instance at point P in Figure 10.2b. Auxiliary power supply sources, for instance batteries, are needed to bias the transistor. Proper biasing is the foremost prerequisite for optimal operation as a signal amplifier. The currents and voltages of the transistor must satisfy the equations:

$$
\begin{aligned}
&I_E = I_0 e^{qV_{BE}/kT} \\
&I_B = I_C/\beta \qquad (10.1) \\
&I_B + I_C = I_E
\end{aligned}
$$

Once biased, the signals (voltages or currents) are superimposed on the bias voltage and current. They are considered to be imposed fluctuations around the bias point. The signal can be applied to the transistor, for instance, by varying the base-emitter voltage or the base current. Whatever variations are made the equations (10.1) still remain valid. When varying one of the transistor currents or voltages, all the other quantities will vary consequently. To avoid non-linearity, the fluctuations are kept relatively small. The description of this small-signal behavior is usually given in terms of the small-signal equations found by differentiating equations (10.1) in the bias point:

$$
\begin{aligned}
&i_e = \frac{qv_{BE}}{kT} I_E = g v_{BE} = \frac{v_{BE}}{r_e} \\
&i_b = i_c/\beta \qquad (10.2) \\
&i_b + i_c = i_e
\end{aligned}
$$

To distinguish between bias quantities and small-signal quantities, the former are written in capitals and the latter in lower-case characters: i_e stands for $\mathrm{d}I_E$ or ΔI_E, v_{be} for $\mathrm{d}V_{BE}$, etc.

The parameter g (A/V or mA/V) is the transconductance of the transistor and represents the sensitivity of the collector current to changes in the base-emitter voltage. Another notation is its reciprocal value, r_e, and the differential emitter resistance (comparable to the differential resistance r_d of a diode).

The transistor's ability to amplify signal power follows from the next few considerations. Assume that the input terminal is the base and the output terminal the collector. Suppose that the input signal is the (change in) base-emitter voltage v_{be}. The collector is furthermore connected to a resistor, R, so that the voltage across R equals $i_c R$. The input signal power is $p_i = i_b v_{be}$, the output signal power is $p_o = i_c^2 R$. The power transfer thus equals $p_o/p_i = i_c^2R/i_bv_{be} = \beta gR$, a value that can easily be much larger than 1.

Equations (10.2) describe the signal behavior, irrespective of the biasing equations (10.1). We use a transistor model based on just the small-signal equations to analyze transistor circuits. An example of such a model, the T-equivalent circuit, is depicted in Figure 10.4a. By applying Kirchhoff's rules it is easy to verify that this model corresponds to the equations (10.2).

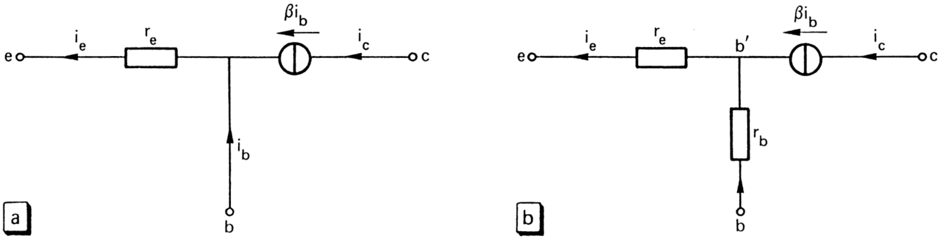


Figure 10.4. (a) A small-signal model of a bipolar transistor, (b) an extended model accounting for the base resistance.

The model can be extended to give a more precise description of the transistor. For example, the model depicted in Figure 10.4b accounts for the base resistance r_b. It is the resistance of the silicon between the base terminal and the internal base contact denoted as b'. The corresponding circuit equations are: $i_er_e = v_{b'e}$ and $i_br_b = v_{b'b}$.

It is also possible to add capacitances to the model in order to also make it valid for high signal frequencies. The next section illustrates how transistor models can be used in the analysis of some basic electronic circuits with bipolar transistors.

10.2 Circuits with bipolar transistors

In the preceding section we saw that a transistor can only operate properly when it is biased correctly and that signals are merely fluctuations around the bias point. In this section we shall examine a number of basic transistor circuits by analyzing both biasing and small-signal behavior.

10.2.1 Voltage-to-current converter

Figure 10.5 shows how to use a bipolar transistor for a voltage-to-current converter: the input voltage V_i is converted to an output current I_o.

First we shall look at the circuit's bias (when the input signal voltage is zero). Two auxiliary voltage sources V^+ (positive) and V^- (negative) are responsible for the biasing. The value of V^+ is such that the base-collector voltage remains positive (to maintain the reverse biased junction) for all possible values of V_i. The output current I_o of this circuit is identical to the collector current I_C, which is almost equal to the emitter current I_E: $I_E = (V_E - V^-)/R_E$. As $V_B = V_i = 0$, the emitter voltage V_E is about –0.6 V (a forward biased diode junction).

When the input voltage V_i changes, emitter voltage V_E consequently changes as well (the difference V_{BE} remains almost unchanged); hence I_E varies and so does I_o.

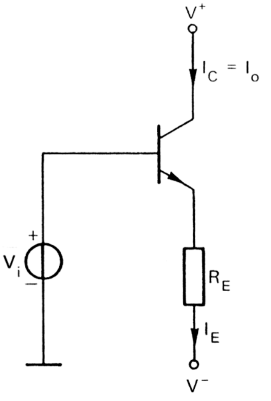


Figure 10.5. A voltage-to-current converter.

Example 10.1
Suppose that $V^+ = 10$ *V;* $V^- = -10$ *V and* $R_E = 4.7$ *kΩ. At zero input voltage,* V_E *is about –0.6 V, so* $I_E = (-0.6 + 10)/4700 = 2$ *mA. This is also the output current* I_o*. With an input voltage of* $V_i = +1$ *V,* V_E *becomes* $1 - 0.6 = 0.4$ *V, and so* $I_E = 10.4/4700 = 2.2$ *mA. For* $V_i = -1$ *V,* $V_E = -1.6$ *V and* $I_E = 8.4/4700 = 1.8$ *mA. The voltage-to-current transfer of this circuit is apparently about* $1/R_E = 0.21$ *mA/V.*

Up until now we have assumed that there is a constant base-emitter voltage of 0.6 V. However, as the current through the transistor changes, so the base-emitter voltage also changes a little. To estimate the significance of this effect, a more precise analysis needs to be made. To that end, the transistor given in Figure 10.5 is replaced by the model given in Figure 10.4a, thus resulting in the circuit of Figure 10.6a.

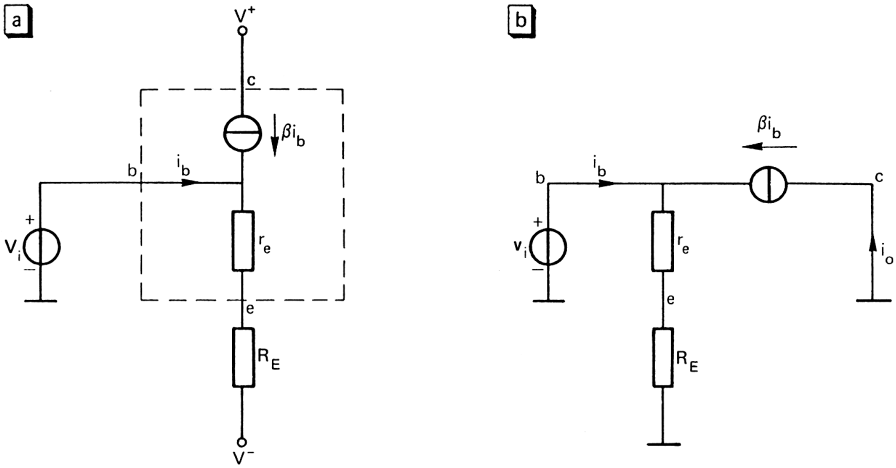


Figure 10.6. (a) A model of the voltage-to-current converter given in Figure 10.5 where the transistor is replaced by the model given in Figure 10.4a, (b) the corresponding small-signal model. As the auxiliary power supply voltages are constant they have zero value in the small-signal model.

The voltages V^+ and V^- of the power sources are constant and have a very low internal resistance: the fluctuations of these voltages is zero and so they can be connected to ground in the small-signal model. The signal model is redrawn in Figure 10.6b. The transistor equations (10.2) are still valid. Using these equations, the output (signal) current i_o can be calculated as a function of the input signal voltage v_i, for instance:

$$v_i = (i_b + \beta i_b)(r_e + R_E) \tag{10.3}$$

$$i_o = i_c = \beta i_b \tag{10.4}$$

from which it follows that:

$$\frac{i_o}{v_i} = \frac{\beta}{(1+\beta)(r_e + R_E)} \tag{10.5}$$

This expression accounts for a finite current gain β and a non-zero emitter differential resistance. For $\beta \to \infty$ and $r_e \to 0$, the transfer is $1/R_E$, a value found earlier.

Example 10.2
Suppose that the transistor in Example 10.1 has a current gain of $\beta = 100$; r_e is 12.5 Ω (because $I_E = 2$ mA). The transfer is $2.10 \cdot 10^{-4}$ A/V, which is almost equal to the approximated value $1/R_E = 2.13 \cdot 10^{-4}$.

Apparently, in order to quickly analyze the circuit, we may consider V_{BE} to be constant (0.6 V), β infinite and r_e small compared to R_E.

10.2.2 The voltage amplifier stage with base-current bias

The base of the transistor in the amplifier circuit shown in Figure 10.7 serves as the input terminal, the output signal is identical to the collector voltage.

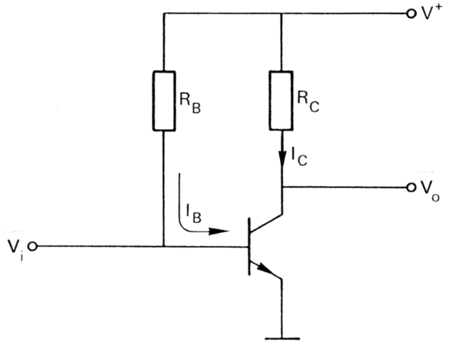


Figure 10.7. A voltage amplifier stage with base current biasing; the bias currents are I_C and I_B.

The bias is created by V^+ and a resistor R_B. The emitter is connected to ground so that $V_E = 0$ and $V_B = 0.6$ V. The base current (which flows through R_B) equals

$(V^+ - 0.6)/R_B$. The collector current is β times the base current. The output bias voltage (V_C) is $V^+ - I_C R_C$.

Example 10.3

Given: $V^+ = +15$ V; $\beta = 100$. Find the value for R_B for which the bias collector current is 1 mA and find the proper value of R_C.

The base current should be 1 mA/100 = 10 μA: $I_B = (15 - 0.6)/R_B = 10^{-5}$, or $R_B = 1.4$ MΩ. The output voltage cannot exceed V^+ and, to prevent the base-collector from being forward biased, it may not drop below V_i. To obtain a maximum range for output variations, set V_o halfway along the outermost values at about +7.5 V. This is achieved for $R_C = (15 - 7.5)/1 \cdot 10^{-3} = 7.5$ kΩ.

The input and output terminals have the emitter in common. Such a circuit is called a common-emitter circuit or CE-circuit.

If this amplifier's input were connected directly to a voltage source the bias point would change dramatically. The same happens when connecting a load to the output (for instance a resistance to ground). To prevent this happening, capacitors are connected in series with the input and the output terminals (coupling capacitors), see C_1 and C_2 in Figure 10.8. The capacitance of these components is so great that they act as a short-circuit for signals.

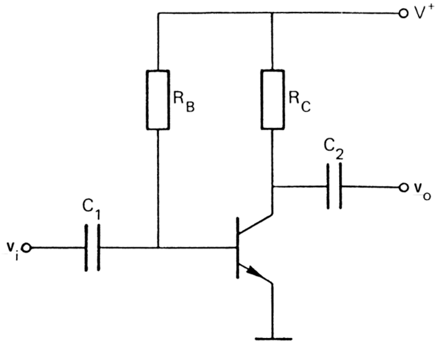


Figure 10.8. The voltage amplifier from figure 10.7, extended with capacitors to connect source and load without affecting the bias.

To calculate the voltage transfer of this amplifier we must make a model of the circuit (Figure 10.9a). This time we shall use the model given in Figure 10.4b to account for the internal base resistance. To simplify the analysis we make the following assumptions: both C_1 and C_2 may be regarded as short-circuits for the signal frequencies, the source resistance $R_g = 0$ (ideal voltage source) and $R_L >> R_C$ (ignoring the load). The model is reduced to resemble that shown in Figure 10.9b.

Voltage transfer $A_v = v_o/v_i$ is established, for instance, from the equations below:

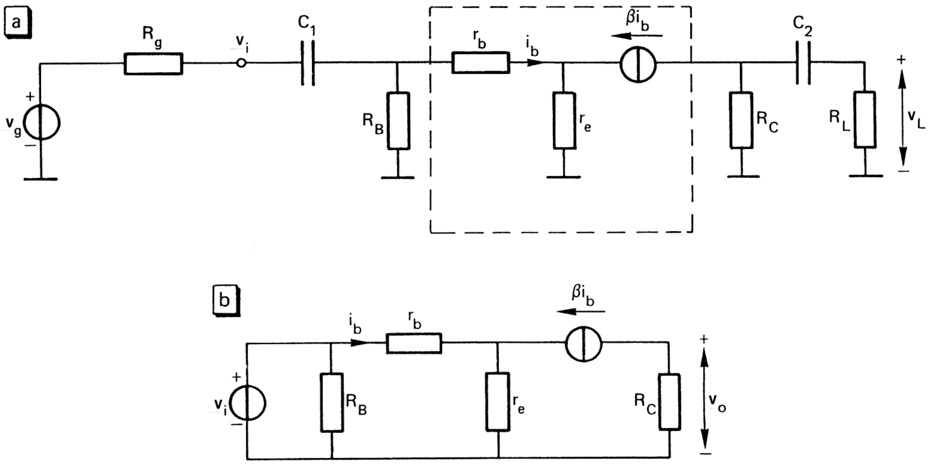


Figure 10.9. (a) Model of the circuit from figure 10.8; (b) simplified model: $R_g = 0$ (hence $v_i = v_g$), $R_L >> R_C$ (hence $v_L = v_o$) and the couple capacitors behave as short-circuits.

$$v_o = -\beta i_b R_C \tag{10.6}$$

$$v_i = i_b r_b + \left(i_b + \beta i_b\right) r_e \tag{10.7}$$

so:

$$A = \frac{v_o}{v_i} = -\frac{\beta R_C}{r_b + (1+\beta) r_e} \tag{10.8}$$

For most transistors, $\beta >> 1$ and $r_b << \beta r_e$, hence $A \approx -R_C/r_e$. The minus sign is of significance because when input voltage increases, collector current will increase and the collector voltage (output) will decrease (in relation to the bias point).

If we want to take into account the effect of the load resistance R_L in the Figure 10.9b model then resistance R_C will simply be replaced by $R_C//R_L$ (the two resistors being in parallel). The voltage gain then becomes roughly $-(R_C//R_L)/r_e$.

In most cases the approximations of the previous analysis are allowed. The discrete resistance values (E12 series, see Section 7.1.1), their tolerances and the tolerances of the transistor parameters make it pointless to carry out analyses that are any more precise.

The biasing of the transistor depends entirely on parameter β. Unfortunately, β varies from transistor to transistor (even if the type is the same) and is temperature-dependent. The biasing (and via r_e also the voltage gain) is therefore not stable or reproducible. To obtain more accurate voltage transfer, irrespective of the transistor parameters, other biasing methods have to be used, as illustrated in the examples given in the following sections.

10.2.3 The voltage amplifier stage with a base-voltage bias

At the amplifier stage shown in Figure 10.10, the base voltage is fixed using voltage divider circuit R_1-R_2 across the supply voltage V^+. A resistor R_E is inserted between the emitter terminal and ground.

The voltage across R_E is $V_B - 0.6$ V, thus resulting in a collector current $I_C \approx I_E = (V_B - 0.6)/R_E$. The input and output voltages are coupled by coupling capacitors to prevent the bias from being affected by source and load circuits.

Further analysis of the amplifier comprises the steps: biasing, signal voltage transfer, input resistance and output resistance.

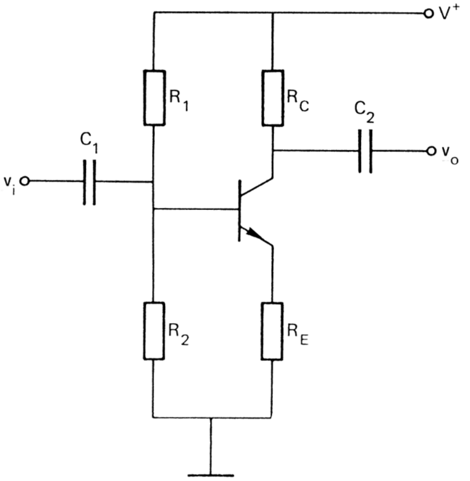


Figure 10.10. A voltage amplifier circuit with base-voltage biasing. The collector bias current is almost independent of the transistor parameters.

Biasing

The resistance values R_1 and R_2 dictate the base voltage: $V_B = V^+R_2/(R_1 + R_2)$, where the bias current is ignored. (This is only permitted when the resistance values of the voltage divider are not too high, the current through R_2 must be large compared to I_B.) As $V_{BE} \approx 0.6$ V, $V_E = V_B - 0.6$ V. The current through R_E is V_E/R_E and equals the collector current. The (bias) output voltage is fixed at $V^+ - I_CR_C$, which should always be larger than V_B.

Small-signal voltage gain

The voltage gain is calculated using a small-signal model of the circuit (Figure 10.11). The coupling capacitors are viewed as short-circuits.

The voltage gain is found using, for instance, the equations:

$$v_i = i_b r_b + (i_b + \beta i_b)(r_e + R_E) \tag{10.9}$$

$$v_o = -\beta i_b R_C \tag{10.10}$$

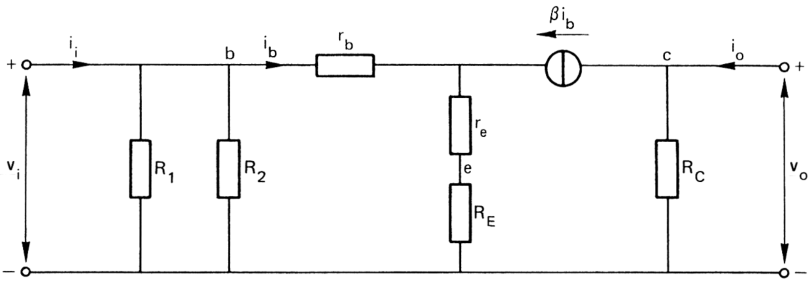


Figure 10.11. A model of the voltage amplifier circuit from figure 10.10.

(without load, $i_o = 0$),
so that

$$A = \frac{v_o}{v_i} = \frac{-\beta R_C}{r_b + (1+\beta)(r_e + R_E)} \approx \frac{-R_C}{r_e + R_E} \approx \frac{-R_C}{R_E} \qquad (10.11)$$

Input resistance

First, for simplicity's sake, we take $R_1//R_2 = R_p$. The elimination of i_o from the next equations:

$$v_i = i_b r_b + (i_b + \beta i_b)(r_e + R_E) \qquad (10.12)$$

$$v_i = (i_i - i_b)R_p \qquad (10.13)$$

results in:

$$r_i = \frac{v_i}{i_i} = R_p \frac{r_b + (1+\beta)(r_e + R_E)}{R_p + r_b + (1+\beta)(r_e + R_E)} = R_p \, // \{r_b + (1+\beta)(r_e + R_E)\} \qquad (10.14)$$

The input resistance of the circuit is equal to a resistance value of $r_b + (1 + \beta)(r_e + R_E) \approx \beta R_E$ in parallel to both biasing resistors which means that: $r_i = \beta R_E // R_1 // R_2$.

Output resistance

The output resistance (at short-circuited input terminals) is found from the equations:

$$v_i = i_b r_b + (i_b + \beta i_b)(r_e + R_E) = 0 \qquad (10.15)$$

$$i_o = \beta i_b + v_0 / R_C \qquad (10.16)$$

From the last equation it follows that $i_b = 0$, so $r_o = v_o/i_o = R_C$, which is simply the collector resistance value.

Resistor R_E has a favorable effect on the biasing which, by then, is almost independent of the transistor parameters. However, the gain factor is reduced to $-R_C/(R_E + r_e)$ compared to $-R_C/r_e$ when R_E is zero. To combine stable bias point with high voltage gain, R_E is decoupled by a capacitor C_E which is in parallel to R_E (Figure 10.12a). This capacitor does not affect the (DC) bias of the circuit. With AC signals it behaves like a short-circuit. The voltage transfer and the input and output impedances of this new circuit can easily be calculated by supplanting R_E in the foregoing expressions with $Z_E = R_E/(1 + j\omega R_E C_E)$. The (complex) transfer function then becomes:

$$\begin{aligned} H = \frac{V_o}{V_i} &= \frac{-R_C}{r_e + R_E/(1 + j\omega R_E C_E)} \\ &= \frac{-R_C}{r_e + R_E} \cdot \frac{1 + j\omega R_E C_E}{1 + j\omega R_E C_E r_e/(r_e + R_E)} \\ &\approx \frac{-R_C}{R_E} \cdot \frac{1 + j\omega R_E C_E}{1 + j\omega r_e C_E} \end{aligned}$$

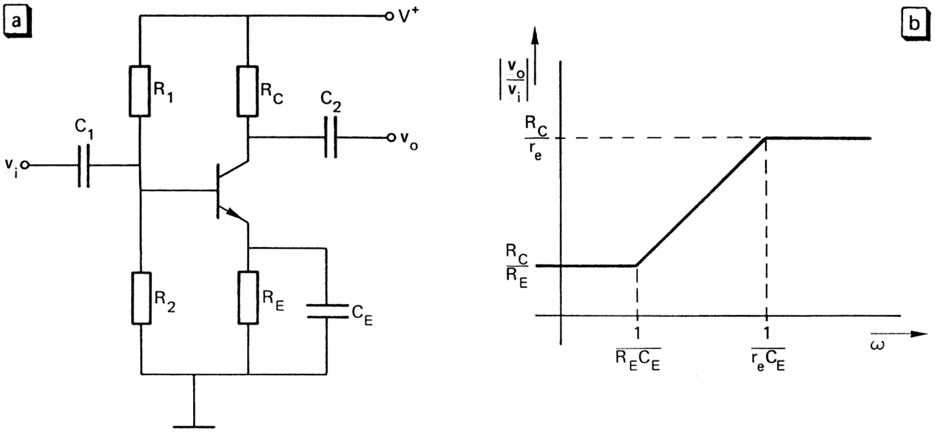


Figure 10.12. (a) A voltage amplifier stage with decoupled emitter resistance; (b) corresponding amplitude transfer characteristic.

For signals with a high frequency ($\omega >> 1/r_e C_E$) the voltage transfer is about $V_o/V_i = -R_C/r_e$ which is the desired high value. Figure 10.12b shows the amplitude transfer as a function of the frequency.

Example 10.4

The given power supply is +15 V. Find proper values for the components in Figure 10.12a so that the voltage gain is $A = -100$, for signal frequencies from 60 Hz.
Step 1: the bias. Let I_E be, for instance, 0.5 mA; this means that $r_e = 1/40I_E = 50\ \Omega$ and $R_C = A \cdot r_e = 100 \times 50 = 5\ k\Omega$. The voltage drop across R_C is $I_C R_C \approx I_E R_C = 2.5$ V so the collector voltage is $15 - 2.5 = 12.5$ V. The voltage may fluctuate around this value but it will never drop below the base voltage so make V_B equal to, for instance, 7.5 V, so that $V_E = 6.9$ V and $R_E = (6.9/0.5) \times 10^{-3} = 13.8\ k\Omega$. The resistances R_1 and

R_2 must be equal in order to get $V_B = 7.5$ V. Take 100 kΩ which is high enough for a reasonable input resistance and should, in fact, be as high as possible but low enough to ignore the influence of the base current at the bias point.
Step 2: the choice of C_E. The lowest frequency f_L must satisfy the inequality $2\pi f_L r_e C_E >> 1$, so $C_E >> 1/2\pi \times 60 \times 50 \approx 53$ μF.

10.2.4 The emitter follower

We noticed that the voltage between the base and the emitter is almost constant. When the base voltage varies, the emitter voltage varies by the same amount. This property is used in a circuit that is denoted as an emitter follower (Figure 10.13). The base is the circuit's input while the emitter acts as output. Although the voltage transfer of an emitter follower is only 1 it does have other useful properties. What exactly these properties are becomes clear when we calculate the input and output resistance of an emitter follower.

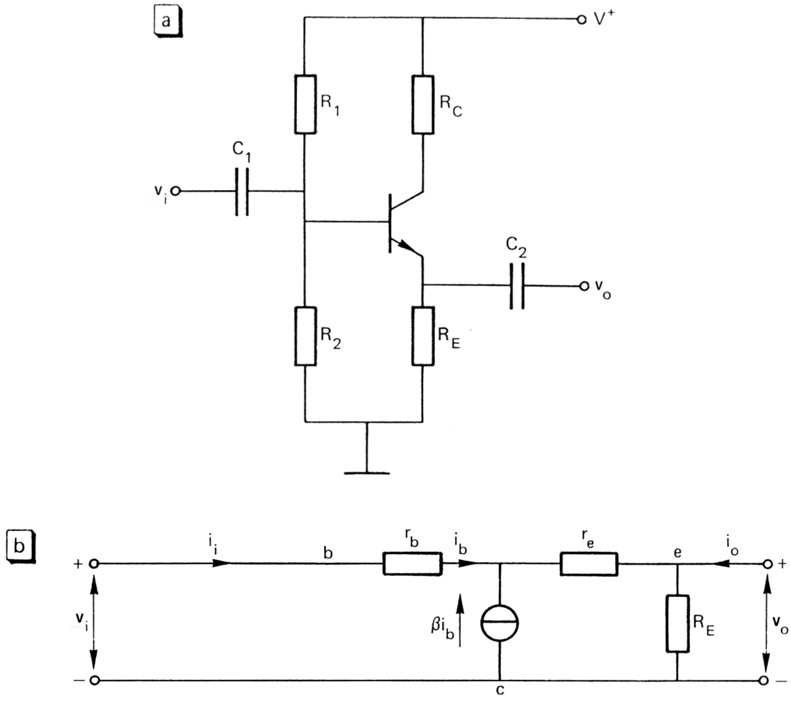


Figure 10.13. (a) An emitter follower; the voltage gain is almost 1;
(b) a model of the emitter follower.

In the model (Figure 10.13b) the bias resistors R_1 and R_2 are omitted to simplify the calculations. Since they are in parallel to the input, they do not affect voltage transfer and output impedance.

Voltage transfer (without load resistor, $i_o = 0$)
The elimination of i_b from the two equations:

$$v_i = i_b r_b + (i_b + \beta i_b)(r_e + R_E) \tag{10.17}$$

$$v_o = (i_b + \beta i_b)R_E \tag{10.18}$$

results in:

$$\frac{v_o}{v_i} = \frac{(1+\beta)R_E}{r_b + (1+\beta)(r_e + R_E)} \approx \frac{R_E}{r_e + R_E} \tag{10.19}$$

For R_E which is large compared to r_e, the transfer is almost 1.

Input resistance (when unloaded, $i_o = 0$)
The elimination of i_b from the next equations:

$$v_i = i_b r_b + (i_b + \beta i_b)(r_e + R_E) \tag{10.20}$$

$$i_i = i_b \tag{10.21}$$

results in:

$$r_i = \frac{v_i}{i_i} = r_b + (1+\beta)(r_e + R_E) \approx \beta(r_e + R_E) \tag{10.22}$$

The resistances R_1 and R_2 are input in parallel so the total input resistance is $r_i = \beta(r_e + R_E)//R_1//R_2$.

Output resistance (at short-circuited input, $v_i = 0$)
The elimination of i_b from the equations:

$$0 = i_b r_b + (i_b + \beta i_b)r_e + v_o \tag{10.23}$$

$$v_o = (i_b + \beta i_b + i_o)R_E \tag{10.24}$$

finally results in:

$$r_o = \frac{v_o}{i_o} = \frac{R_E\left(r_e + \dfrac{r_b}{1+\beta}\right)}{R_E + r_e + \dfrac{r_b}{1+\beta}} = R_E \,//\left(r_e + \frac{r_b}{1+\beta}\right) \approx r_e \tag{10.25}$$

From this analysis it emerges that an emitter follower has a high input resistance (roughly β times R_E) and a low output resistance (about r_e). The emitter follower is therefore suitable as a buffer amplifier stage between two voltage transfer circuits and for minimizing load effects.

Example 10.5
Let the bias current of an emitter follower be 1 mA. Other transistor parameters are: $\beta = 200$ *and* $r_b = 100\ \Omega$. R_E *is 10 k*Ω. *In this situation the voltage transfer is approximately* 104/(25 + 10^4) ≈ 0.9975, *the input resistance is* $200 \times (10^4 + 25) \approx 2$ MΩ *and the output resistance is 25* Ω. *The factor* $r_b/(1 + \beta)$ *can be ignored compared to* r_e.

10.2.5 The differential amplifier stage

A major disadvantage of all the coupling capacitor circuits discussed so far is that they cannot cope with DC signals. The coupling capacitors separate the bias quantities (DC) from the signal quantities (AC). In order to prevent the input voltage source and the load becoming part of the bias, coupling capacitors should be applied. Even if we succeed in solving this problem there is still the other problem of temperature. We know that at constant (bias) current the base-emitter voltage varies according to temperature (–2.5 mV/K). Such slow changes are not distinguished from gradually changing input signals.

Most of these problems are solved using the circuit type shown in Figure 10.14. The basic idea is to compensate the temperature sensitivity of the transistor by one second in an identical transistor. First we shall discuss the biasing of this amplifier stage before going on to study its signal behavior.

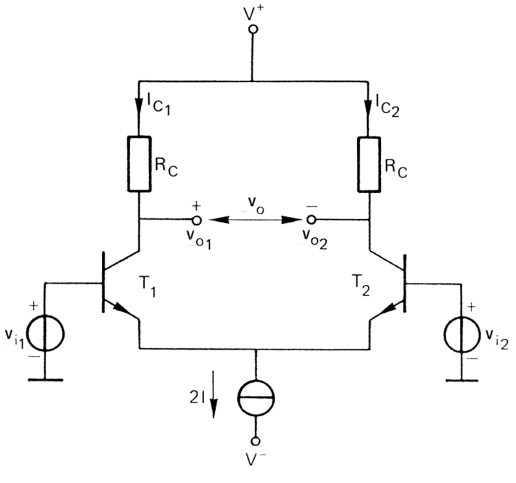


Figure 10.14. A differential amplifier stage: only the difference between v_{i1} and v_{i2} is amplified.

Bias

The base terminals of the transistors T_1 and T_2 act as input terminals for the amplifier. With respect to biasing, both base contacts are at ground potential. The bias current is supplied by a current source with a current of $2I$ like, for instance, the circuit given in Figure 10.5 with a fixed input voltage. The base-emitter voltages of both transistors are equal so their currents: $I_{C1} = I_{C2}$ are the same. The collector voltage is set by V^+ and R_C: $V_C = V^+ - I_C R_C$. Naturally this voltage must be higher than the highest occurring base voltage.

Signal properties

As long as $v_{i1} = v_{i2}$, the base-emitter voltages of both transistors remain equal, their collector currents therefore remain I each. The amplifier is not sensitive to equal input signals or common mode signals. However, for $v_{i1} = -v_{i2}$ the collector currents of T_1 and T_2 will change but their sum remains $2I$. For positive v_{i1} the collector current through T_1 increases by v_{i1}/r_e, whereas the collector current of T_2 decreases with the same amount: $-v_{i2}/r_e$. The output voltages (i.e. the collector voltages) of T_1 and T_2 are thus $v_{o1} = -v_{i1}R_C/r_e$ and $v_{o2} = +v_{i2}R_C/r_e$, respectively. Having a differential voltage $v_d = v_{i1} - v_{i2}$ between both inputs gives rise to a differential output $v_o = v_{o1} - v_{o2} = -R_C v_d/r_e$. The transfer for differential voltages is $-R_C/r_e$, just as at the normal CE-amplifier stage. The transfer for common input signals appears to be zero. Due to the asymmetry of the circuit, that is to say, the unequal transistor parameters, the common mode transfer may differ slightly from zero. The ratio between the differential transfer and the common mode transfer is the common mode rejection ratio or CMRR of the differential amplifier (Section 1.2).

This circuit, also known as the long-tailed pair, forms the basic circuit for almost all types of operational amplifiers (see further Chapter 12).

SUMMARY

The properties of bipolar transistors

- There are two types of bipolar transistors: pnp- and npn-transistors.
- The three parts of a bipolar transistor are: the emitter, the base and the collector. The currents through the corresponding terminals are denoted as: I_E, I_B and I_C.
- An essential prerequisite for the proper operation of a bipolar transistor is the very narrow width of its base region.
- The current gain of a bipolar transistor is $\beta = I_C/I_B$. The typical values of β are between 100 and 300. So, $I_E \approx I_C$.
- The collector current satisfies the equation $I_C = I_o e^{qV_{BE}/kT}$ (and is like that of the pn-diode). If the base-collector junction is reverse biased, I_C will be virtually independent of the collector voltage.
- The bipolar transistor functions as a voltage controlled-current source (from V_{BE} to I_C) or a current amplifier with a current gain of β.
- A bipolar transistor is biased and has fixed DC voltages and currents. Signals are treated like fluctuations around the bias point.
- The change in the collector current, i_c, will only depend on the change in base-emitter voltage: $i_c = g \cdot v_{be}$, where $g = 1/r_e$ is the transconductance of the transistor

and r_e the emitter differential resistance: $r_e = kT/qI_C$. Hence $r_e \approx 25\ \Omega$ at $I_C = 1$ mA (compare with the r_d of a pn-diode).

Circuits with bipolar transistors

- When used as a signal amplifier the base-emitter junction of a bipolar transistor is forward biased and the base-collector junction is reverse biased.
- The analyzing of an electronic circuit with bipolar transistors can be divided into two parts: biasing and small-signal behavior.
- When analyzing the small-signal behavior of a transistor circuit a transistor model is used in which all the fixed voltages, emanating for instance from the power source, are zero.
- To prevent the source and load affecting the bias point they are coupled to the circuit via couple capacitors. This will consequently set a lower limit for the signal frequency.
- The voltage transfer of a CE-stage with emitter resistance R_E and collector resistance R_C is roughly $-R_C/R_E$. With decoupled emitter resistance the transfer becomes approximately $-R_C/r_e$. The low frequency cut-off point is about $\omega = 1/C_E r_e$.
- An emitter follower has a voltage transfer of 1, a high input resistance (of about βR_E) and a low output resistance (of about r_e). The circuit is used as a voltage buffer.
- The first part of a differential amplifier is designed for low offset and drift and for a high common mode rejection ratio. The latter is limited by the asymmetry of the components.

EXERCISES

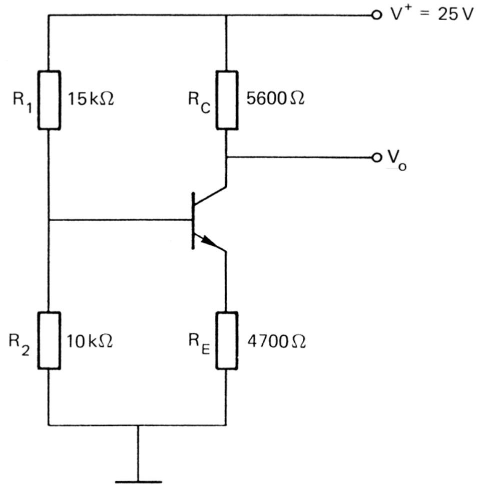
The properties of bipolar transistors

10.1 Give the relationship between the collector current I_C and the base-emitter voltage V_{BE} of a bipolar transistor used in a linear amplifier circuit. Mention the conditions for V_{BE} and V_{BC} for proper operation.

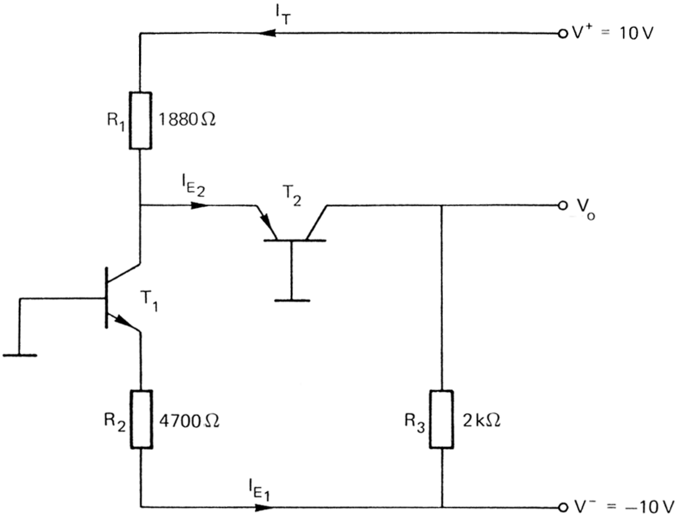
10.2 The current gain of a bipolar transistor is $\beta = 200$. Find the base current and the collector current for an emitter current, I_E, of 0.8 mA.

10.3 What phenomenon is described by the Early effect?

10.4 In the circuit below, $V_{BE} = 0.6$ V. Find the output voltage V_o. The base current can be ignored.


V⁺ = 25 V
R1
15 kΩ
RC
5600 Ω
Vo
R2
10 kΩ
RE
4700 Ω

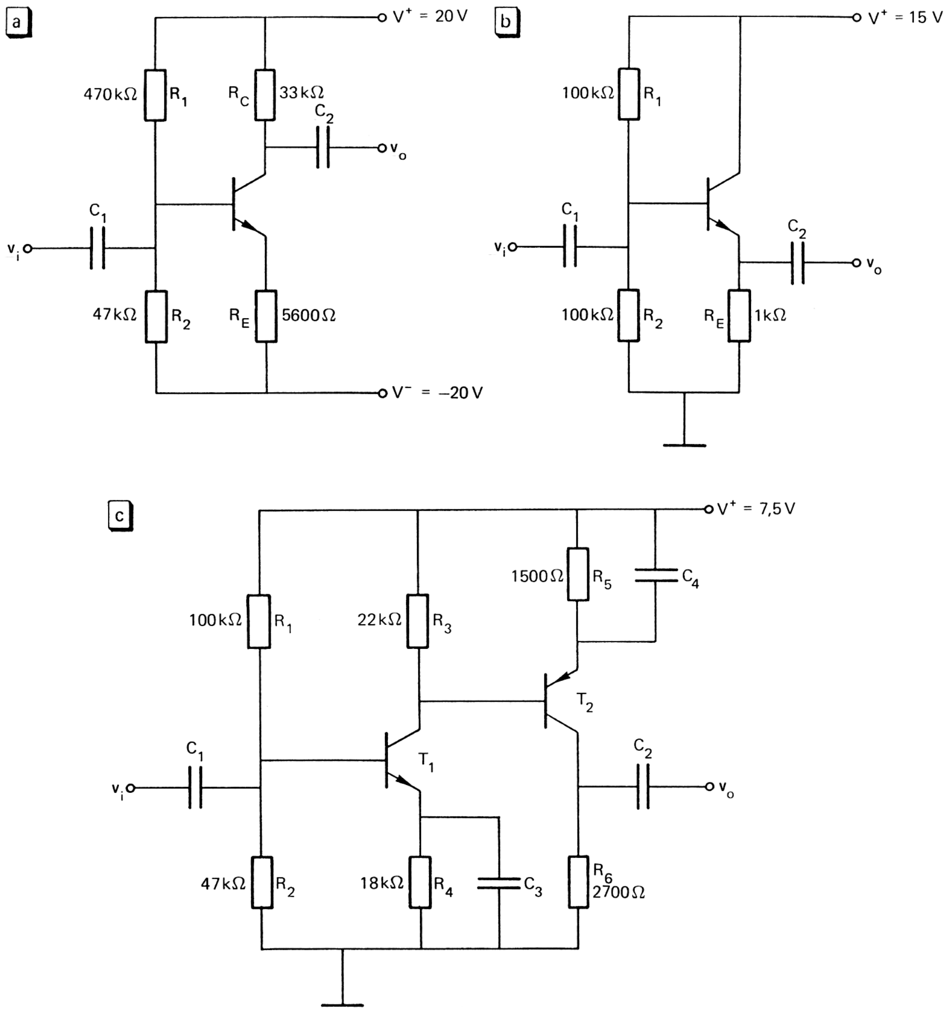
10.5 In the following figure, V_{BE} = 0.6 V for the npn transistor, and $V_{EB} = -V_{BE}$ = 0.6 V for the pnp transistor. Find the output voltage V_o. Clue: first calculate I_{E1}, then I_T and finally I_{E2}.


IT
V⁺ = 10 V
R1
1880 Ω
IE2
T2
Vo
T1
R2
4700 Ω
R3
2 kΩ
IE1
V⁻ = −10 V

Circuits with bipolar transistors

10.6 Take the next three voltage amplifier stages a-c (figure on next page). For all transistors is $\beta = \infty$, r_b = 0 Ω and V_{BE} = 0.6 V (for the pnp: –0.6 V). Calculate the differential emitter resistance of each of the transistors.

10.7 Refer to the previous exercise. Find the voltage transfer of each of those circuits. The coupling capacitors may be seen as short-circuits for signals, furthermore $r_e << R_E$.


a
V⁺ = 20 V
470 kΩ
R1
RC
33 kΩ
C2
vo
C1
vi
47 kΩ
R2
RE
5600 Ω
V⁻ = −20 V
b
V⁺ = 15 V
100 kΩ
R1
C1
vi
C2
vo
100 kΩ
R2
RE
1 kΩ
c
V⁺ = 7,5 V
1500 Ω
R5
C4
100 kΩ
R1
22 kΩ
R3
T2
C1
vi
T1
C2
vo
47 kΩ
R2
18 kΩ
R4
C3
R6
2700 Ω

10.8 The voltage transfer v_o/v_i in exercise 10.6b is not exactly equal to 1, due to $r_e \neq 0$. What is the deviation from 1?

10.9 Due to a finite value of β and the fact that $r_b \neq 0$, the voltage transfer v_o/v_i of circuit 10.6b deviates somewhat from 1. Find that deviation for $\beta = 100$ and $r_b = 100\ \Omega$.

10.10 Calculate the input resistance and the output resistance of circuit 10.6b, taking into account the components R_1, R_2 and the transistor parameters β and r_b: $\beta = 100$ and $r_b = 100\ \Omega$. Approximations up to ±5 % are allowed.

10.11 In the circuit given in Exercise 10.6c the output of the first stage (i.e. the collector of the npn-transistor) is loaded with the input of the second stage (the base of the pnp-transistor). Calculate the voltage transfer reduction due to this load for the whole circuit and compare it to the value found in exercise 10.7. Take $\beta = 100$.

10.12 Calculate the emitter differential resistance r_e of the transistor in Exercise 10.6b, assuming that $\beta = 100$.

11 Field-effect transistors

A field-effect transistor or FET is an active electronic component suited to signal amplification similar to that seen in bipolar transistors. Other possible uses for the field-effect transistor are as a voltage-controlled type of resistance and as an electronic switch. Although the physical mechanisms of an FET essentially differ from those of a bipolar transistor, it is remarkably similar when used as a signal amplifying component: the FET is also biased with DC voltages and the signals are treated like fluctuations around the bias point.

There are two main kinds of FETs: the junction field-effect transistor or JFET and the metal-oxide semiconductor field-effect transistor or MOSFET. We will first explain the operating principles and list the main properties of FETs. In the second part of the chapter certain linear amplifier circuits will be described in which FETs serve as active components.

11.1 The properties of field-effect transistors

The word field-effect alludes to the possibility of being able to affect the conductivity of a semiconductor material by means of an electric field. A JFET utilizes the particular properties of a pn-junction to create voltage-dependent conductivity. In a MOSFET the conductivity is affected by capacitive induction. In both cases, the concentration of free charge carriers (electrons or holes) is controlled by a voltage.

11.1.1 Junction field-effect transistors

In Chapter 9 we saw that the initial width of the depletion layer depends on the doping concentration of the materials. The width can furthermore be modulated by the voltage across the junction. The JFET is based on this latter property. Figure 11.1 provides a schematic representation of the structure of a JFET. It consists of a p-doped or n-doped silicon substrate with a thin layer of the complementary type of silicon. The layer is provided with two contacts called source and drain. The path between these contacts is the channel. If the channel is n-type then the FET will be an n-channel FET, otherwise it is a p-channel FET. In normal operation, the substrate and the channel are electrically separated from each other by a reverse voltage across the junction. The conductance of the channel between source and drain (lateral conductance) depends on the channel dimensions, in particular the effective thickness, that is to say, the part of the layer that contains free charge carriers. The depleted part of the channel does not contribute to conductance. When the reverse voltage across the pn-junction is increased the depletion layer width will increase as well, thus reducing the effective thickness of the channel and its conductance. When a voltage is introduced between the drain and the source, a current will flow along the channel between these contacts. That current can be controlled by the substrate voltage. The connection to the substrate is called the gate electrode or simply the gate.

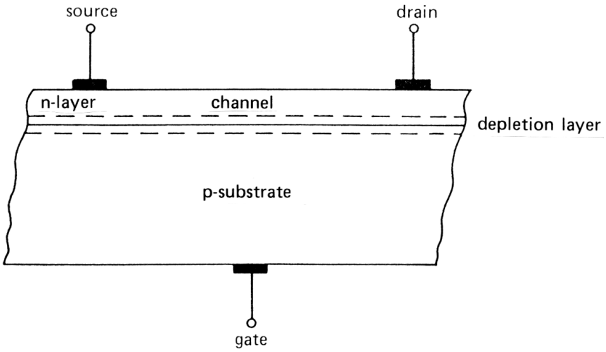


Figure 11.1. Schematic structure of an n-channel JFET.

To obtain high sensitivity or, a substantial change in channel conductance while varying the gate voltage, the channel must be very thin, preferably the same width as a depletion layer (<1 μm). Modern technology makes it possible for such thin channel layers that can be completely depleted to be created. At zero channel conductance, the FET is said to be pinched-off. The gate voltage at which this occurs is termed the pinch-off voltage V_p and that is an important JFET parameter.

Figure 11.2a shows the voltage-current characteristic of the channel for low currents and for various values of the gate-source voltage. Apparently in this region the JFET – the channel resistance – behaves like a voltage-controlled resistance. At pinch-off the resistance is infinite and at $V_{GS} = 0$ the channel resistance has a relatively low value of generally around several 100 Ω.

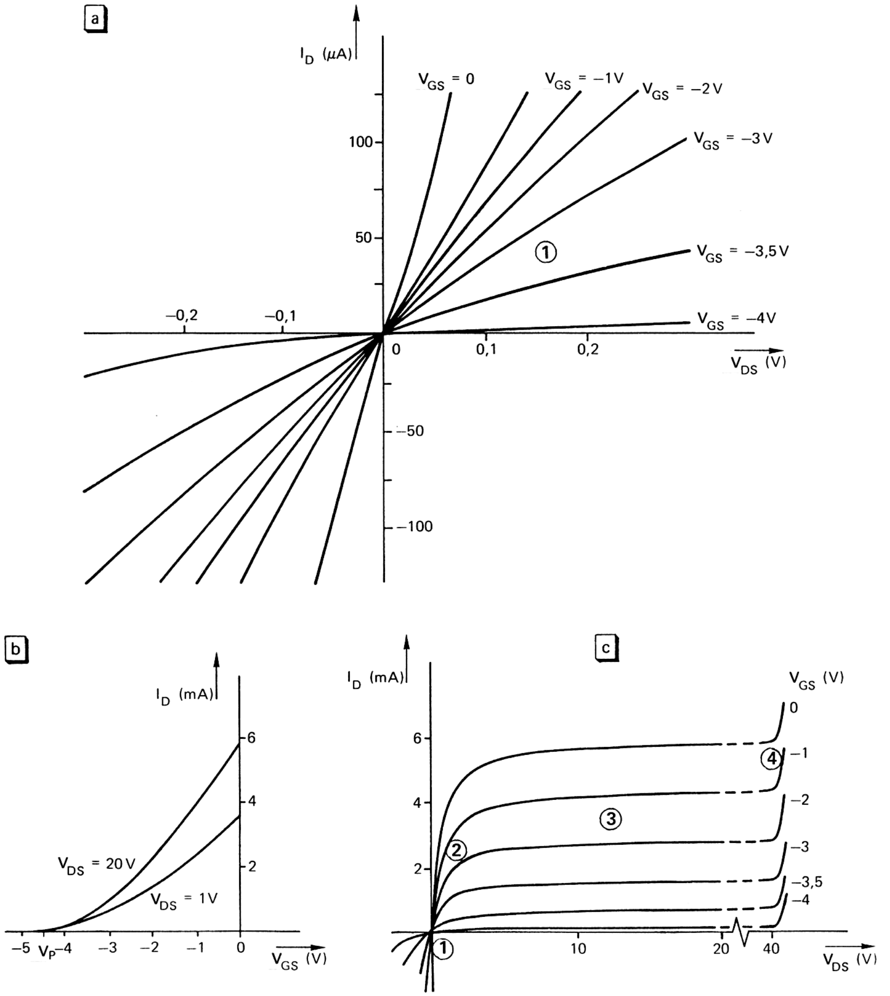


Figure 11.2. (a) The drain current versus the drain-source voltage of a JFET for low current values and for various values of the gate-source voltage, (b) drain current versus gate-source voltage for two values of V_{DS}, (c) drain current versus drain-source voltage for various values of V_{GS}. When drain-source voltage becomes too high breakdown occurs (region 4). Note that the voltages of an FET are usually relative to the source voltage of V_{GS} and V_{DS}.

The depletion layer is determined by the voltage across the junction, in other words, by the voltage between the source and the gate, V_{GS}, and between the drain and the gate, V_{DS}. To explain the other curves given in Figure 11.2 we shall start with $V_{GS} = 0$ and look to see what happens when only the drain voltage is increased. A slight increase in V_{DS} will not affect the depletion layer width, its resistance will remain constant. The channel current from source to drain therefore increases linearly as the drain voltage increases, just as with a normal resistor. This is the curve in Figure 11.2a at $V_{GS} = 0$ and in region 1 of Figure 11.2c. When further increasing the drain voltage the voltage across the junction that is near to the drain contact will gradually increase, as does the width of the depletion layer. The conductance of the channel thus decreases. As the source-gate voltage is still zero, the width of the depletion layer in the vicinity of the source contact does not change. This explains the triangular shape of the depletion region in Figure 11.3a.

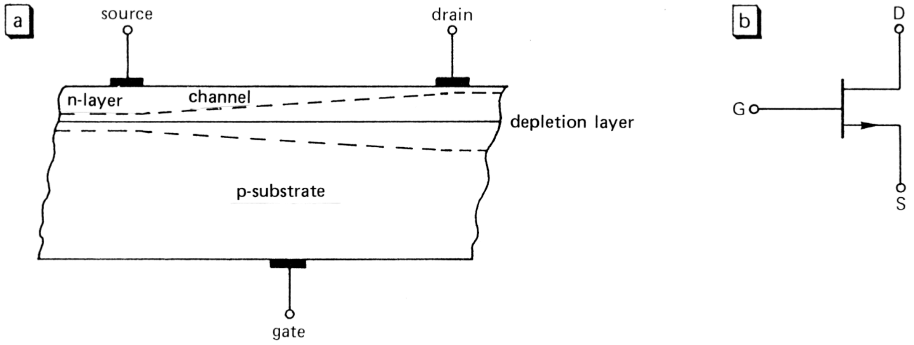


Figure 11.3. (a) The shape of the depletion region for zero gate-source voltage and positive drain-source voltage, (b) the circuit symbol for a JFET.

Due to decreasing conductance, the drain current will increase at a rate that is less than proportional to the drain voltage (region 2 in Figure 11.2c). Finally, when the drain voltage and thus the decreasing channel conductance is further increased, the current will reach saturation level (region 3 in Figure 11.2c, upper curve). By this stage the channel is almost pinched off, the current will flow through a narrow path near the drain and will be prevented from further growing, even at higher drain voltages.

If the process above is started with a negative gate-source voltage, the initial channel width will be smaller in accordance with the higher value of the resistance. The current-voltage curves in Figures 11.2a and c will become less steep when the gate-source voltage is more negative and the drain current will be saturated at a much lower value (region 3, other curves).

In region 3 the drain current is independent of the drain voltage and is determined only by the gate-source voltage. In this region, the JFET behaves like a voltage-controlled current source similar to that of the bipolar transistor. This similarity is reflected in the circuit symbol of the JFET (Figure 11.3b).

Because of the different physical mechanisms when compared to the bipolar transistor the JFET also needs to be differently biased. In the bipolar transistor one junction is forward biased while the other is reverse biased. In the JFET, the channel must always be reverse biased to prevent conduction to the substrate. Both source-gate and drain-gate junctions must therefore be reverse biased. This means that with an n-channel JFET both the source and the drain should be positive with respect to the gate (or for a p-channel just negative). An important spin-off is the very low gate current consisting of only the reverse biased pn-junction's leakage current.

In the saturation region or pinch-off region the theoretical relationship between the gate-source voltage and the drain current is:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_p}\right)^2 \tag{11.1}$$

in which I_{DS} is the current for $V_{GS} = 0$ (see also Figure 11.2b). Just as in the bipolar transistor, the transconductance g of a JFET is defined as the ratio between changes in I_D and V_{GS}:

$$g = \frac{dI_D}{dV_{GS}} = \frac{i_d}{v_{gs}} = \frac{-2}{V_p\sqrt{I_{DS}I_D}} \tag{11.2}$$

So g is proportional to $\sqrt{I_D}$. A typical value for the pinch-off voltage is a few volts, I_{DS} ranges from several mA up to 100 mA. The transconductance of a JFET is, therefore, around 1 mA/V at $I_D = 1$ mA. This is much less than the transconductance of the bipolar transistor at the same bias current. On the other hand, the gate current of the JFET is much lower than the base current, so the current gain with a JFET, I_D/I_G, is very high.

In most cases, the gate current of the JFET can be ignored There is just one equation to describe the small-signal behavior of the JFET:

$$i_d = gv_{gs} \tag{11.3}$$

The model of a JFET is very simple too, see Figure 11.4a. This model can be extended to account for all kinds of deviations from the ideal behavior. For example, the model given in Figure 11.4b accounts for the influence that the drain voltage has on the drain current (compare the Early effect in a bipolar transistor).

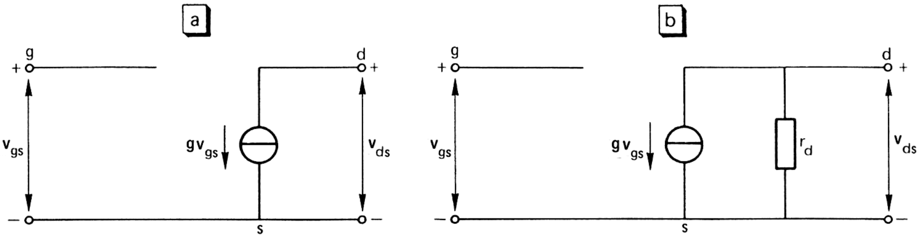


Figure 11.4. (a) A model of a JFET, (b) an extended model taking into account the drain differential resistance.

11.1.2 MOS field-effect transistors

The operation mechanism of a MOSFET differs in at least two ways from that of the JFET. Firstly, this is because the channel conductivity is not controlled by the substrate voltage but rather by the voltage of an isolated electrode connected to the top of the channel. The isolation consists of a very thin layer of silicon dioxide (Figure 11.5a).

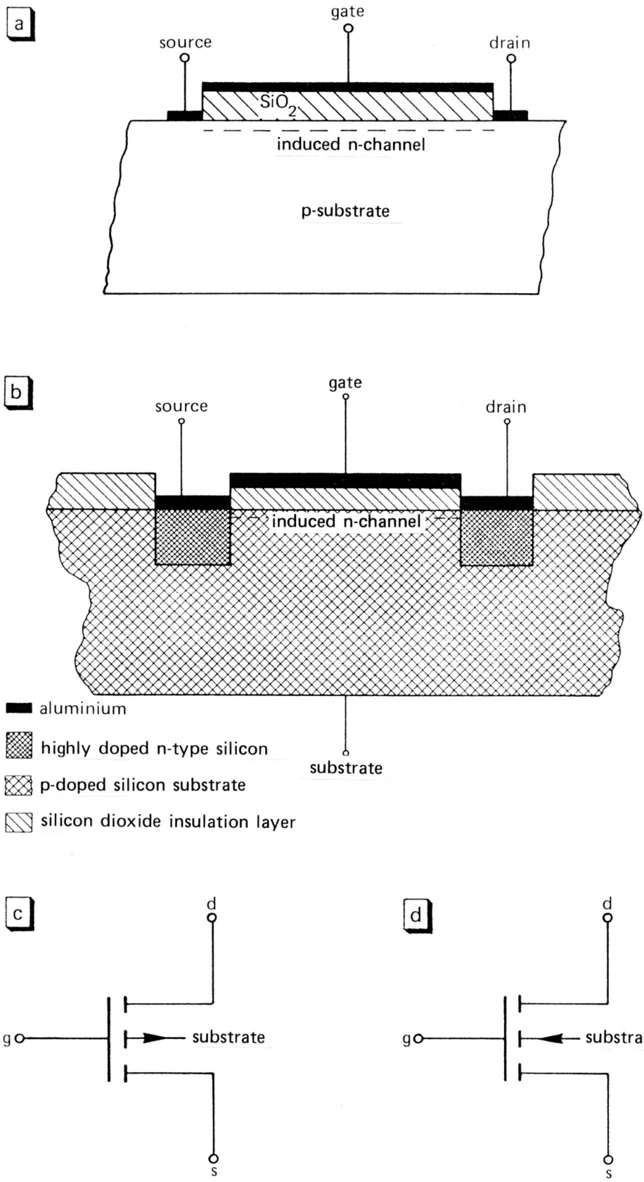

a
gate
source
drain
SiO2
induced n-channel
p-substrate
b
gate
source
drain
induced n-channel
aluminium
highly doped n-type silicon
p-doped silicon substrate
silicon dioxide insulation layer
substrate
c
d
g
substrate
s
d
g
substrate
s

Figure 11.5. (a) the Schematic structure of a MOSFET, (b) the realization of an n-channel MOSFET. The source and drain consist of highly doped n-type silicon with a very low resistance, (c) & (d) circuit symbols for an n-channel and a p-channel MOSFET.

Secondly, the conducting channel is not a deposited layer but an induced layer. On the surface of the silicon crystal the regular structure is broken. The atoms on the surface do not find neighboring atoms to share their valence electrons. A large concentration of holes exists on the surface which is then still electrically neutral and the absent material acts as a donor. Electrons from the adjacent oxide easily recombine with those holes. When that happens, though, the top layer of the silicon becomes negatively charged producing a more or less conductive channel near the surface. The concentration of electrons trapped on the surface of the p-silicon may become so high that the channel starts to behave like an n-type silicon. This is what is called an inversion layer.

This phenomenon is employed when creating a MOSFET. Two contacts, the drain and the source, form the end points of a channel that then consists of such an inversion layer. The structure is completed by a metal gate contact on top of the isolating oxide layer. The concentration of electrons in the channel can be decreased simply by introducing a negative gate voltage. The negative voltage then pulls holes from the p-region towards the surface where they recombine with the electrons thus reducing the channel conductivity. When the gate voltage is sufficiently high, the channel conductance becomes zero and the MOSFET is pinched off, as in a JFET.

Apparently this type of MOSFET conducts at zero gate voltage. It is a normally-on type transistor, like a JFET (a bipolar transistor, by contrast, is normally-off). It is possible to construct MOSFETs with initially low charge carrier concentrations in the induced channel. Only when the gate voltage is sufficiently positive will enough electrons be driven into the channel to make it conductive. This type of transistor is called a normally-off MOSFET. Clearly there are four types of MOSFETs: p and n-channel, both of which are either normally-on or normally-off.

As the gate electrode is fully isolated from the drain and source electrodes by an oxide layer the gate current is extremely small, somewhere in the region of 10 fA. Consequently the current gain of the MOSFET is almost infinite. One disadvantage of the MOSFET is the relatively low breakthrough voltage of the thin oxide layer (10 to 100 V). The gate contact may not be touched because then it will become electrostatically charged.

Another disadvantage is the poor noise behavior. The MOSFET mechanism relies entirely on the surface properties of the crystal, the slightest impurity will affect its operation. Nevertheless, MOSFETs are used extensively in digital integrated circuits, where noise is less of a problem and where most interconnections are made internally, within the package. Furthermore, because of their simple lay-out, MOSFETs can be made very small thus producing high component chip density.

Figure 11.5b provides the schematic structure of a MOSFET. The source and drain contacts are formed on small areas of highly doped n-silicon in a lightly doped p-substrate. This results in a low contact resistance between the external electrodes and the channel material. However, two new pn-junctions are then introduced which go from both contacts to the substrate. To make sure that these junctions do not have any influence they must be kept reverse biased. For that purpose MOSFETs have a special connection, a substrate connection that is different from the gate. By connecting the substrate to the most negative voltage in a circuit, preferably the negative power voltage source, the junctions always stay reverse biased.

11.2 Circuits with field-effect transistors

We are confining all the discussion in this chapter to the area of junction field-effect transistors. All the bipolar circuits introduced in Section 10.2 can be operated with JFETs as well. It is only the biasing that differs, both the gate-source and the gate-drain junctions must be reverse biased. The gate of an n-channel FET should always be negative with respect to the source and the drain. To numerically determine the bias, the characteristics of the JFET need to be known, for instance those given in Figures 11.2b and c.

11.2.1 Voltage-to-current converter

In the voltage-to-current converter of Figure 11.4 the bipolar transistor is replaced by a JFET (Figure 11.6). Let us just suppose that the characteristics of Figure 11.2 apply to this type of JFET. A bias current of, for instance, 2 mA will occur for $V_{GS} = -2$ V (the JFET operates in region 3). At $V_i = 0$ (bias condition), the source voltage is +2 V (V_{GS} is negative). The value of R_S must be such that $I_D R_S = V_S - V^-$, or $R_S = (2 - V^-)/2\times10^{-3}$. It is possible to choose $V^- = 0$ (no negative power source required) so that $R_S = 1$ kΩ. Evidently V^+ has to be sufficiently positive.

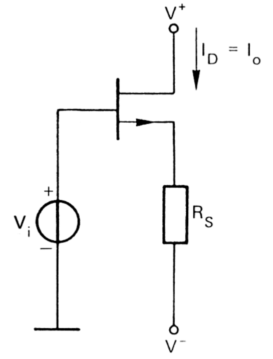


Figure 11.6. A voltage-to-current converter with a JFET.

The transfer of the converter is found in a similar way to the circuit of the bipolar transistor. Using the model given in Figure 11.4a what we find for the transfer is:

$$\frac{i_o}{v_i} = \frac{1}{R_S + \dfrac{1}{g}} \tag{11.4}$$

The rather low value of g (compared to that of the bipolar transistor) does not allow it to be ignored.

11.2.2 The voltage amplifier stage

Both JFET junctions must be reverse biased and the biasing is quite easy (Figure 11.7). The couple capacitors make the bias independent of the source and load circuits. Resistor R_G is inserted to assure that the very low gate current amount flows to ground. If this resistor were left out, the gate current would not be able to flow anywhere and would hamper the proper bias. R_G can be very large, for instance 10 MΩ. The gate voltage, $I_G R_G$, almost stays at zero.

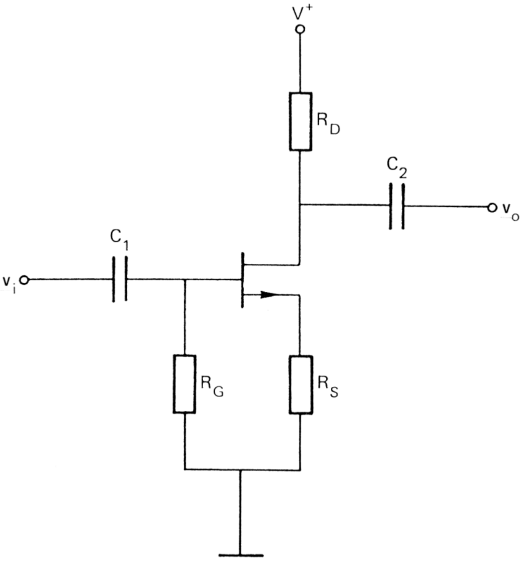


Figure 11.7. A voltage amplifier stage with JFET.

Example 11.1

Suppose we want a bias current of 4 mA. From the I_D-V_{GS} characteristic (Figure 11.2b) it follows that there is a gate-source bias voltage of –1 V so R_S = 250 Ω. The drain-source voltage should be biased in the region of constant drain current, so it should be at least 5 V (Figure 11.2c). The bias point is fixed half way through the available range, between 6 V and V^+. For a power supply voltage of V^+ = 20 V a proper bias value of V_D is +13 V, which is achieved by having a drain resistance of $(20 – 13)/4\times10^{-3}$ = 1.75 kΩ. At a specified maximum gate current of, for instance, 200 pA the gate resistance can be as high as 10 MΩ: the gate voltage is only $I_G R_G$ = 1 mV, a value that barely influences the bias point.

The voltage transfer, the input resistance and the output resistance can be calculated with the model given in Figure 11.8, which uses the transistor model of Figure 11.4.

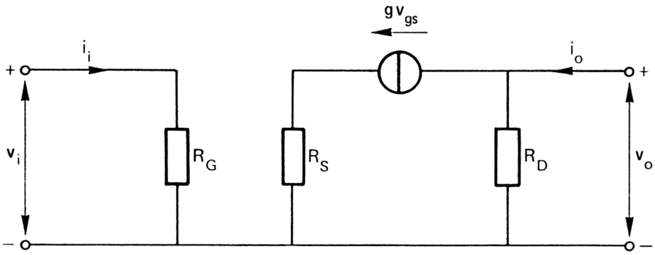


Figure 11.8. A model of the amplifier circuit from Figure 11.7, the couple capacitors are viewed as short-circuits.

voltage transfer (for $i_o = 0$):

$$\frac{v_o}{v_i} = \frac{-gR_D}{1 + gR_S} \tag{11.5}$$

input resistance:

$$r_i = R_G \tag{11.6}$$

output resistance:

$$r_o = R_D \tag{11.7}$$

Example 11.2

From the characteristic in Figure 11.2b it follows that there is a transconductance of about 2 mA/V. Using the bias conditions given in Example 11.1, we find that there is a value of $v_o/v_i = -3.5/1.5 = -2.3$, for the voltage transfer there is an input resistance of $r_i = 10\ M\Omega$ and the output resistance is $r_o = 1.75\ k\Omega$. The voltage gain can be somewhat increased by decoupling resistor R_S, thus resulting in a voltage transfer of $-gR_D = -3.5$. The rather high output resistance requires a sufficiently high input resistance at the next stage.

11.2.3 The source follower

The circuit seen in Figure 11.6 acts as a source follower when it is not the drain current but rather the source voltage that is taken as the output (Figure 11.9a). The source follower has the same function as the emitter follower. It has a voltage transfer of 1 and high input resistance and low output resistance.

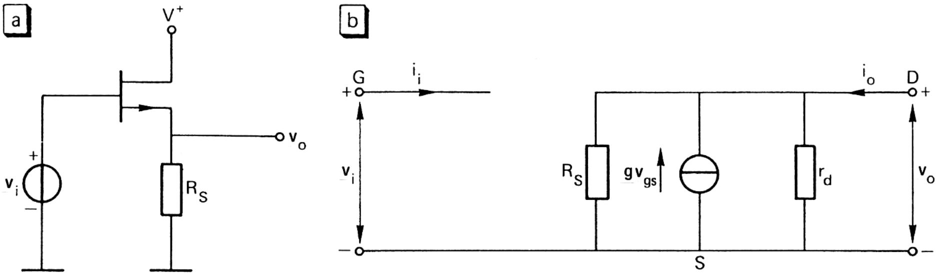


Figure 11.9. (a) A source follower with a voltage gain of 1, (b) a model of the source follower.

The properties of the source follower are analyzed on the basis of a model of Figure 11.9b. This model also accounts for the effect of a finite value of r_d.

voltage transfer (for $i_o = 0$):
The elimination of v_g and v_s from the equations

$$v_o = g v_{gs}(R_S // r_d) \tag{11.8}$$

$$v_g = v_i \tag{11.9}$$

$$v_s = v_o \tag{11.10}$$

results in:

$$v_o = g(v_i - v_o)(R_S // r_d) \tag{11.11}$$

and so:

$$\frac{v_o}{v_i} = \frac{g(R_S // r_d)}{1 + g(R_S // r_d)} \tag{11.12}$$

input resistance:
infinite, because $i_i = 0$.

output resistance (at $v_i = 0$):
the output resistance follows from the equation

$$v_o = (i_o + g v_{gs})(R_S // r_d) = (i_o - g v_o)(R_S // r_d) \tag{11.13}$$

hence

$$r_o = \frac{v_o}{i_o} = \frac{R_S // r_d}{1 + g(R_S // r_d)} \tag{11.14}$$

Example 11.3
Suppose that the transconductance of the JFET applied to the circuit is 2 mA/V and r_d = 100 kΩ. To obtain a voltage transfer of 1, the term $g(R_S//r_d)$ must be as high as possible. Take, for instance, R_S = 10 kΩ. This results in $g(R_S//r_d) \approx 18.2$, so v_o/v_i = 0.95 and r_o = 474 Ω.

SUMMARY

The properties of field-effect transistors

- A junction field-effect transistor (JFET) consists of an n or p-channel the thickness of which, and thus the conductance of which, depends on the width of the depletion layer which, in turn, is affected by the reverse voltage.
- The three terminals of a JFET are drain, source and gate. The channel is located between the source and drain.

- The gate-source voltage for which channel conductance is just zero is the pinch-off voltage V_p.
- The JFET behaves like a voltage-controlled current source. The relation between changes in the drain current and the gate voltage is: $i_d = gv_{gs}$, where g is the transconductance of the FET.
- To operate in analog signal processing circuits the FET is biased with DC voltages and currents. The signals are fluctuations around the bias point.
- The gate current of a JFET is the very small leakage current of the reverse biased channel-substrate junction. This gate current is much smaller than the base current of a bipolar transistor.
- The gate of a MOSFET is isolated from the channel by a thin oxide layer. The gate current of a MOSFET is extremely small.
- A MOSFET has a separate substrate terminal which (for n-channel types) must be biased on the most negative voltage in the circuit.

Circuits with field-effect transistors

- A junction field-effect transistor (JFET) is biased so that the source-gate and the drain-gate junctions are reverse biased.
- The analysis of an electronic circuit with JFETs can be split up into two parts: comprising biasing and small-signal behavior.
- To analyze the small-signal behavior of a transistor circuit a transistor model is used in which all fixed voltages, for instance from the power source, are zero.
- To prevent the source and load affecting the bias, parts of the circuit can be coupled by coupling capacitors. This will, as a result, set a lower limit for the signal frequency.
- With JFETs, similar circuits can be created to those made with bipolar transistors. The main differences are:
  - the very small gate current so that a very high input resistance can be achieved,
  - smaller transconductance which means that the voltage gain is reduced,
  - that the influence of the drain voltage on the drain current is not negligible, it is represented by the internal drain resistance r_d.

EXERCISES

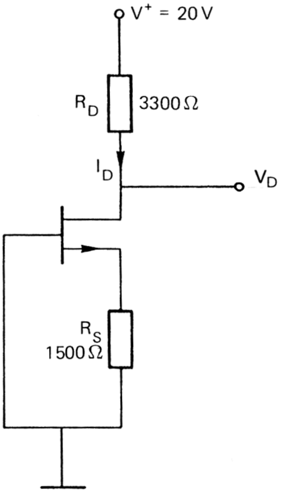
The properties of field-effect transistors

11.1 What is the pinch-off voltage of a JFET?

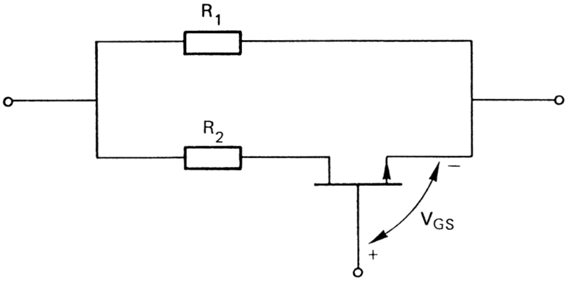
11.2 The region around a pn-junction is called the depletion layer, why?

11.3 Compare the JFET and the MOSFET with respect to their gate currents.

11.4 Find the drain voltage V_D and the drain current I_D in the circuit given below, for $V_{GS} = -4$ V.



11.5 The channel resistance of a certain JFET type varies from 800 Ω (at $V_{GS} = 0$) to ∞ ($V_{GS} < V_p$). With the network shown below we want to design a voltage-controlled resistance that varies from 1 kΩ to 10 kΩ for the same range of V_{GS}. Find the values for R_2 and R_2.

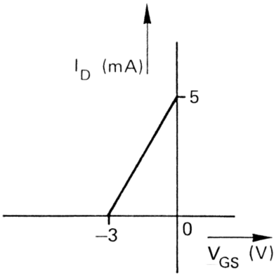


11.6 Explain the need for a substrate MOSFET terminal.

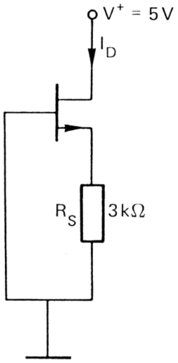
11.7 Field-effect transistors are also called unipolar transistors as opposed to the bipolar transistors described in Chapter 10. Explain these terminology discrepancies.

Circuits with field-effect transistors

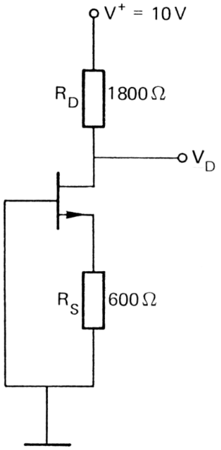
11.8 The I_D-V_{GS} characteristic of a JFET is illustrated in the next figure. Find the expression for $I_D = f(V_{GS})$.



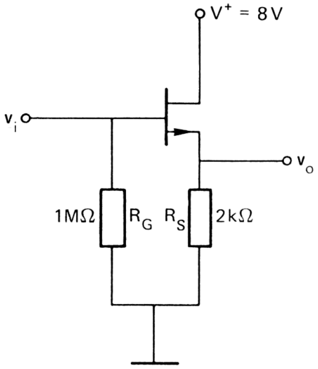
11.9 Calculate the drain current of the circuit given in the next figure. The characteristics in the preceding exercise also apply to this JFET.


V⁺ = 5V
I_D
R_S
3kΩ

11.10 Calculate V_D in the circuit given below. The FET has the characteristics described in Exercise 11.8.


V⁺ = 10V
R_D
1800Ω
V_D
R_S
600Ω

11.11 Make a model of the source follower given below and find the voltage transfer, the input resistance and the output resistance under the conditions: $g = 2$ mA/V and $i_g = 0$.

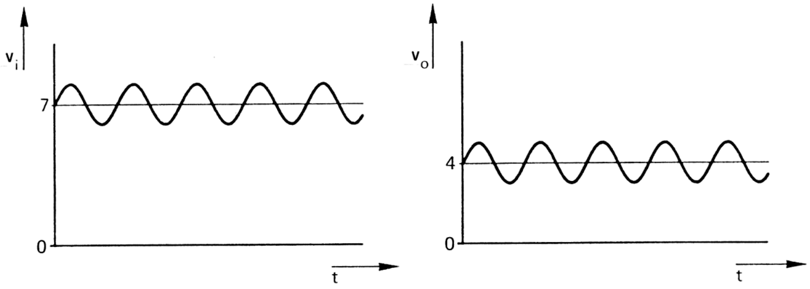

V⁺ = 8V
v_i
v_o
1MΩ
R_G
R_S
2kΩ

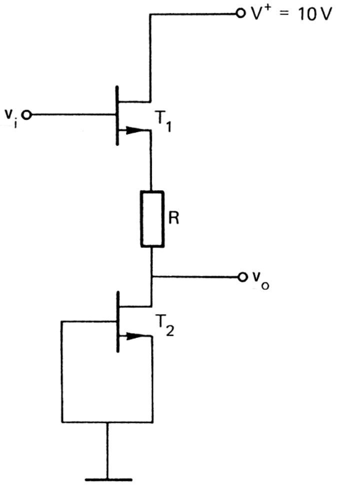
11.12 We are given an AC voltage with a DC component of 7 V (see Figure below). We want to transfer this signal to an average level of 4 V without affecting the AC part. This is achieved by introducing the next circuit (a level shifter).

a. Explain how the circuit operates.

b. For both JFETs: I_D (at $V_{GS} = 0$) = 2 mA. Calculate R.

c. What is the maximum allowed peak value of v_i?





12 Operational amplifiers

There are a number of drawbacks attached to amplifiers with only one transistor, particularly where the matters of: gain, input impedance, output impedance and offset are concerned. The way to obtain better amplifiers is by establishing appropriate combinations of various circuits. Such configurations can soon become highly complex, indeed their design has become something of a specialist field. Fortunately, since the technology of integrated circuits has become so advanced, the users of electronic systems no longer need to do all their own designing. The operational amplifier makes it possible for a whole range of high quality processing circuits to be quite simply configured. The basic concepts underlying such designs are given in the first part of this chapter. Ultimately, even operational amplifiers have their shortcomings as is evidenced by the circuits designed. The second part of this chapter addresses the issue of design analysis and puts forward various solutions to reduce amplifier errors..

12.1 Amplifier circuits with ideal operational amplifiers

An operational amplifier is essentially a differential amplifier (Section 1.2) that has very high voltage gain A (10^4 to 10^6), a high CMRR ($>10^4$) and a very low input current. These kinds of amplifiers have a large number of circuit components like transistors, resistors and, quite often, several capacitors. They do not have inductors. All the components are integrated into a single silicon crystal or chip that is mounted on a metal or plastic encapsulation. Figure 12.1a portrays the circuit symbol, and Figures 12.1b and c depict two common types in various encapsulations, together with the pin layout and pin functions. The two inputs are known as the inverting input (indicated by the minus sign) and the non-inverting input (denoted by the plus sign).

The most important imperfections of an operational amplifier are the offset voltage (Section 1.2) and the input bias currents (i.e. small but noticeable DC currents flowing through the input terminals and caused by the base or gate currents in the input transistors, Sections 10.1 and 11.1).

The offset voltage $V_{i,off}$ can be accounted for by placing a voltage source in series with one of the two input terminals. The two bias currents $I_{bias,1}$ and $I_{bias,2}$ can be modeled on the basis of two current sources but an easier way to account for them would be by placing arrows next to the input leads (Figure 12.1d).

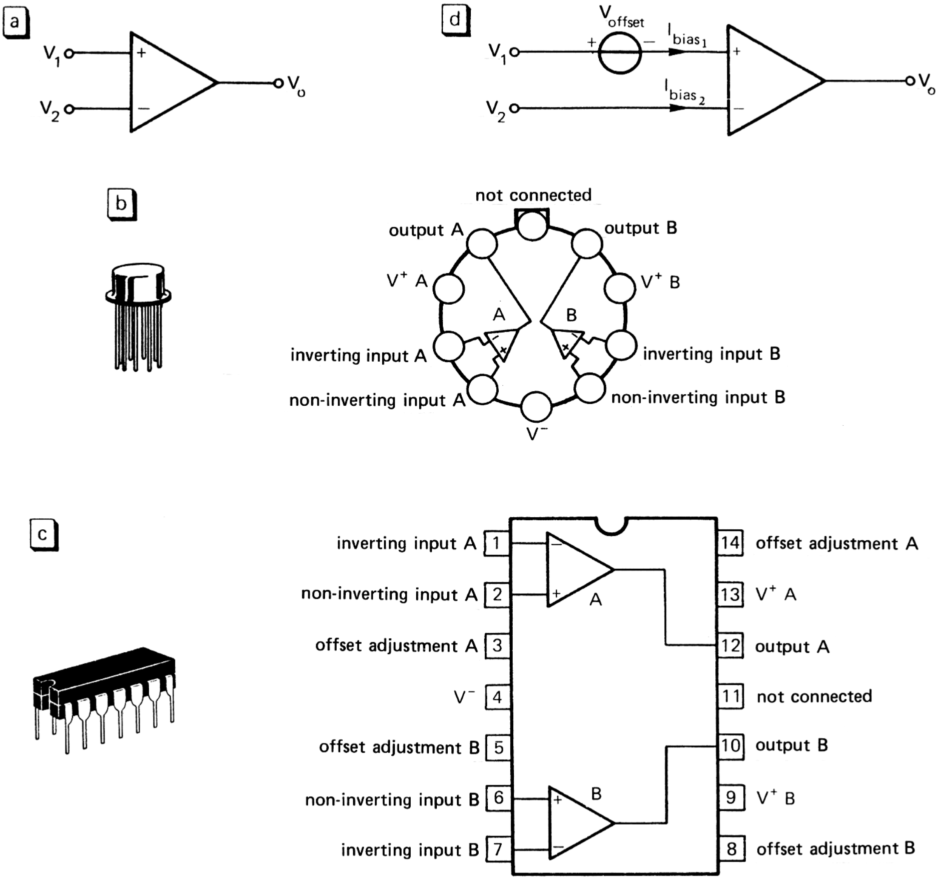

a
V_1
V_2
+
−
V_o
d
V_{offset}
+
−
I_{bias_1}
I_{bias_2}
V_1
V_2
+
−
V_o
b
not connected
output A
output B
V⁺ A
V⁺ B
A
B
inverting input A
inverting input B
non-inverting input A
non-inverting input B
V⁻
c
inverting input A 1
non-inverting input A 2
offset adjustment A 3
V⁻ 4
offset adjustment B 5
non-inverting input B 6
inverting input B 7
14 offset adjustment A
13 V⁺ A
12 output A
11 not connected
10 output B
9 V⁺ B
8 offset adjustment B
A
B

Figure 12.1. (a) The circuit symbol of an operational amplifier, $V_o = A(V_1 - V_2)$, (b) & (c) two types of dual operational amplifiers with corresponding pin layout and pin function indication, (d) the modeling of the input offset voltage and the two bias currents.

The most important imperfections of an operational amplifier are the offset voltage (Section 1.2) and the input bias currents (i.e. small but noticeable DC currents flowing through the input terminals and caused by the base or gate currents in the input transistors, Sections 10.1 and 11.1).

The offset voltage $V_{i,off}$ can be accounted for by placing a voltage source in series with one of the two input terminals. The two bias currents $I_{bias,1}$ and $I_{bias,2}$ can be modeled on the basis of two current sources but an easier way to account for them would be by placing arrows next to the input leads (Figure 12.1d).

The operational amplifier is nearly always used in combination with a feedback network. An open amplifier will become saturated because of the high voltage gain and the non-zero input offset voltage. Its output, which is either maximally positive or maximally negative, is limited by the power supply voltage.

In this section we assume that the behavior of the operational amplifier is ideal. This means that there will be: infinite gain, bandwidth, CMRR and input impedance as well as zero input offset voltage, bias currents and output impedance. Manufacturers will specify these and a lot of other parameters. Appendix B.2.1 gives an example of the complete specifications for a widely-used type of operational amplifier.

Just how circuits with ideal operational amplifiers are designed will be illustrated using a number of frequently implemented circuits.

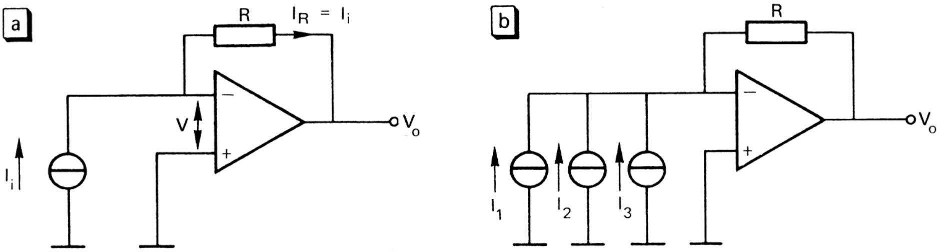


Figure 12.2. (a) A converter for the conversion of a current I_i into a voltage V_o, (b) an extended circuit for the summation of currents: $V_o = -R(I_1 + I_2 + I_3)$.

12.1.1 Current-to-voltage converters

In Figure 12.2a we see a current-to-voltage converter circuit diagram with an operational amplifier. Let us imagine that V is the voltage across the operational amplifier's input terminals (here that is also the negative input voltage because the positive input is grounded). The output voltage is $V_o = A \cdot V$. The input current for an ideal amplifier is zero, so in line with Kirchhoff's currents' rule, input current I_i must flow through feedback resistance R. Kirchhoff's rule for voltages states that $-V + I_iR + V_o = 0$, therefore $V = I_iR - V/A$ or: $V = I_iR/(1 + A)$. The gain A is very high (∞ in ideal cases), so V is almost zero. A key property of an ideal operational amplifier is that the voltage difference between the two input terminals is zero. The output voltage of the converter is simply $V_o = -I_iR$.

At first sight, the combination of infinite gain and finite output voltage seems strange but this is all down to the feedback because the output is partly fed back into the input. If V_o were for some reason to grow, then the input voltage V would increase as well, however, the current through R remains constant and thus also the voltage across it which tends to reduce the output change. The result is an equilibrium in which V is exactly zero. In practice, A is finite but very large so the voltage between the two input terminals is not zero but very small or negligible compared to other voltages in the circuit.

In the circuit given in Figure 12.2 the non-inverting input is ground connected which means that the voltage of the inverting input also has zero potential: it is said to be virtually grounded. The voltage is zero but it is not a real ground, there is in fact no ground connection.

The two properties of the operational amplifier, the zero input voltage and the zero input currents, very much simplify the analysis of such circuits. They make it possible to directly calculate the transfer of the circuit shown in Figure 12.2a. The procedure is as follows: I_i must flow entirely through R and, as $V = 0$, the output voltage is determined via $0 = I_i R + V_o$.

When a second current source is connected to the converter input that current will flow entirely through R. The same holds for a third source, and so on (see Figure 12.2b). This circuit acts as an adder for currents. The output voltage satisfies $V_o = -R(I_1 + I_2 + I_3)$.

12.1.2 Inverting voltage amplifiers

The inverting voltage amplifier that can be seen in Figure 12.3a may be viewed as a combination of a voltage-to-current converter and a current-to-voltage converter.

The input voltage V_i is converted into a current with resistor R_1. The inverting input of the operational amplifier is virtually grounded, its potential is zero, so $I = V_i/R_1$. This current flows entirely through R_2, so $V_o = -R_2 I = -(R_2/R_1)V_i$. The amplifier circuit gain $-R_2/R_1$, is determined by the ratio of two external resistances and is independent of the operational amplifier parameters. The accuracy and stability are only determined by the quality of the resistors used, not by that of the amplifier. In the special case of two equal resistances: $R_1 = R_2$, the output is $V_o = -V_i$; the gain is -1 and the voltage is inverted.

This circuit can be extended to include the summation of voltages (Figure 12.3b). The output voltage of this circuit is $V_o = -(V_1/R_1 + V_2/R_2 + V_3/R_3)R_4$ which is a weighted summation of the three input voltages.

The input resistance of the inverting amplifier to be seen in Figure 12.2a is just equal to R_1. The resistance values are determined by, on the one hand, the specified transfer and on the other by the lowest input resistance. Ideally the output resistance is zero.

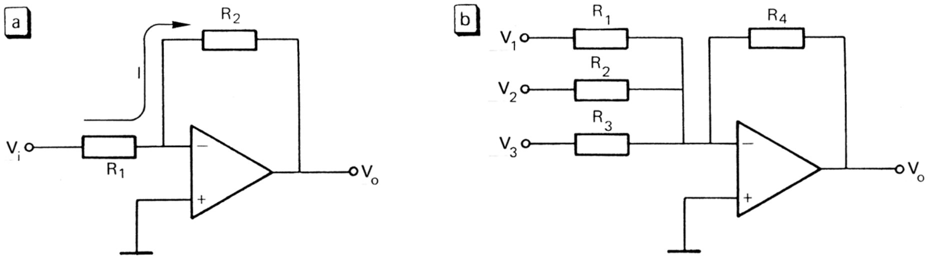


Figure 12.3. An inverting voltage amplifier (a) with single input, (b) as an adder, with multiple input.

Example 12.1

Let the required gain of an amplifier be –50 and the input resistance at least 13 kΩ. If $R_1 = 15$ *kΩ, then* $R_2 = 750$ *kΩ.*

12.1.3 Non-inverting voltage amplifiers

The operational amplifier's output terminal given in Figure 12.4a is directly connected to the negative input terminal (this is called unity feedback). The input voltage is connected to the positive input terminal.

The voltage transfer of this circuit can be found in the following way: the voltages at the positive and negative input terminals are equal, their difference is zero, so $V_o = V_i$, the output follows the input voltage which means that the gain is +1. The circuit has a very high input impedance equivalent to the operational amplifier and a very low output impedance which is the same as that of the operational amplifier. The circuit acts as a buffer amplifier.

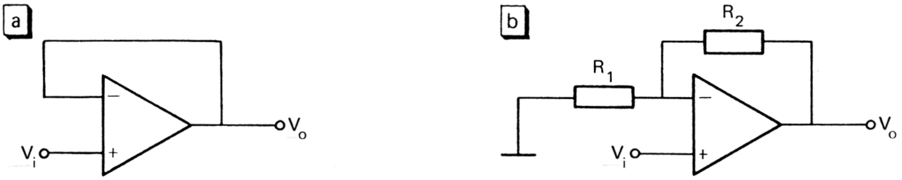


Figure 12.4. Non-inverting amplifiers, (a) with voltage transfer 1 (buffer amplifier), (b) with arbitrary voltage transfer >1, set at R_1 and R_2.

In the circuit given in Figure 12.4b only part of the output is fed into the input. Again, the voltages of both operational amplifier input terminals are equal, so the voltage of the negative input terminal is V_i. This voltage is also equal to the output of the voltage divider, it is $R_1/(R_1 + R_2)$ of the output. This means that:

$$\frac{V_o}{V_i} = \frac{R_1 + R_2}{R_1} \tag{12.1}$$

Again, the gain is set by the ratio of just two resistances.

12.1.4 Differential amplifiers

If an inverting and a non-inverting amplifier are combined it then becomes possible to subtract voltages. Figure 12.5 shows the most simple configuration.

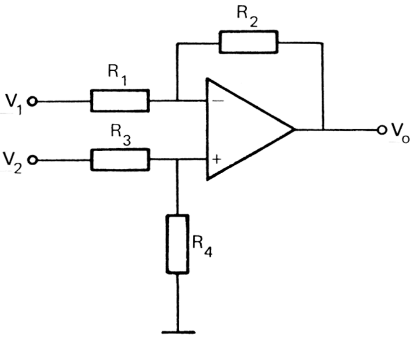


Figure 12.5. A differential amplifier circuit.

We use the principle of superposition (the circuit is linear) to determine the transfer. First we simply determine the output voltage due to V_1, the other input voltage is zero (grounded). The plus-terminal must therefore also be zero (no current flows through R_3 and R_4). This situation is identical to the configuration seen in Figure 12.3a where $V_o = -(R_2/R_1)V_i$. After that we have to establish the output voltage due only to V_2 where V_1 is taken to be zero. The voltage at the plus-terminal of the operational amplifier becomes the output of the voltage divider with R_3 and R_4, hence $V_2 \cdot R_4/(R_3 + R_4)$. When V_1 is grounded the circuit is identical to the configuration of Figure 12.4b where the transfer is $(R_1 + R_2)/R_1$ and the input voltage is the same as that mentioned above. Hence:

$$V_o = \frac{R_1 + R_2}{R_1}\frac{R_4}{R_3 + R_4}V_i \tag{12.2}$$

The total output voltage is found by adding the contributions made by V_1 and V_2:

$$V_o = -\frac{R_2}{R_1}V_1 + \frac{R_1 + R_2}{R_1}\frac{R_4}{R_3 + R_4}V_2 \tag{12.3}$$

Let us now consider a special case: $R_1 = R_3$ and $R_2 = R_4$. When the four resistors are in this state the output is

$$V_o = \frac{R_2}{R_1}\left(V_2 - V_1\right) \tag{12.4}$$

which is the amplified difference of both input voltages. The transfer of this differential amplifier is determined by the four resistances R_1 to R_4 regardless of the properties of the operational amplifier.

When we compare the properties of this new differential amplifier with the operational amplifier itself we see that the input resistance is much lower (input resistances are now R_1 and $R_3 + R_4$, for the two inputs), that its gain is much lower (though more stable) and that the CMRR is much lower (due to the tolerances of the resistors).

The CMRR is defined as the ratio of the differential mode gain and the common-mode gain (Section 1.2). We will calculate the CMRR for the new configuration.

Suppose that the resistances have relative inaccuracy ε_i (i = 1 to 4), so $R_i = R_i^*(1 + \varepsilon_i)$, where R_i^* is the nominal value of R_i. Suppose also that $\varepsilon_i \ll 1$, and $R_1^* = R_3^*$ and $R_2^* = R_4^*$ (the condition for a differential amplifier). A pure differential mode voltage can be given as $V_d = V_1 - V_2$ or $V_1 = \frac{1}{2}V_d$ and $V_2 = -\frac{1}{2}V_d$. The transfer for this input voltage is $A_d = V_o/V_d \approx -R_2/R_1$.

A pure common mode voltage is written as V_c or $V_1 = V_2 = V_c$; the transfer for such signals is

$$A_c = \frac{V_o}{V_c} = \frac{R_1R_4 - R_2R_3}{R_1(R_3 + R_4)} \approx \frac{R_2^*}{R_1^* + R_2^*}(\varepsilon_1 + \varepsilon_4 - \varepsilon_2 - \varepsilon_3) \tag{12.5}$$

As the sign of the relative errors ε is not known, we take the modulus $|\varepsilon|$ to find the worst case CMRR:

$$CMRR = \frac{A_d}{A_c} = \frac{1 + R_2/R_1}{|\varepsilon_1| + |\varepsilon_2| + |\varepsilon_3| + |\varepsilon_4|} \tag{12.6}$$

Example 12.2

A differential amplifier is built in the way illustrated in Figure 12.5 and the resistors have a specified inaccuracy of ±0.5%. The gain of such an amplifier is –100. To find the CMRR, we conclude that $A_d = R_2/R_1 = 100$*, so the guaranteed rejection ratio is* $101/(4\times 0.005) = 5050$*.*

12.1.5 Instrumentation amplifiers

A major disadvantage of the differential amplifier type shown in Figure 12.5 is its rather low input resistance. This can be rectified by connecting buffer amplifiers (Figure 12.4a) to inputs. In that way the transfer will not change but the input resistance will be the same as the buffers and therefore very high. Another drawback of the differential amplifier is its low CMRR as determined by the tolerances of the resistors. When two buffers are inserted the CMRR will be even lower. If the buffers have slightly different gains the common mode signal at their inputs will result in a small differential signal where the two buffers have their output. In turn, that will be amplified and made into a differential mode signal by the differential amplifier. To circumvent this difficulty we use the arrangement given in Figure 12.6.

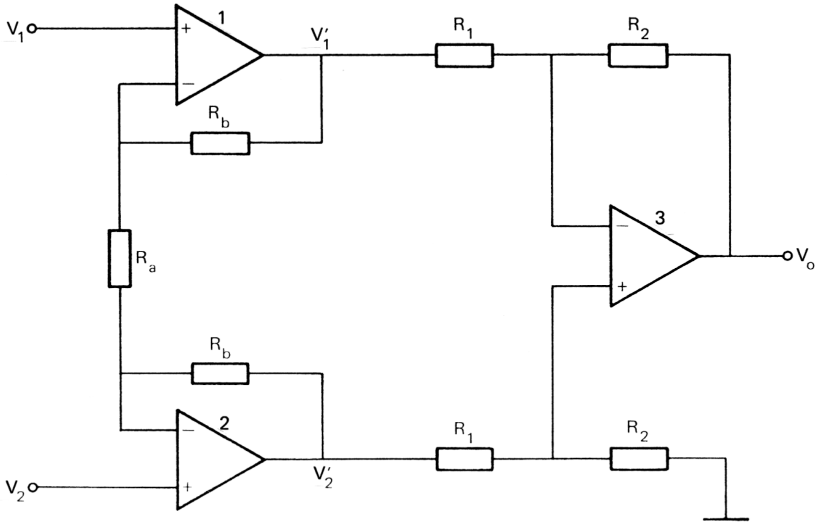


Figure 12.6. An instrumentation amplifier with high input resistance and high CMRR.

The transfer and the CMRR are determined as follows. First we consider a pure common mode input signal: $V_1 = V_2 = V_c$ that is connected to the non-inverting inputs of operational amplifiers 1 and 2. The inverting input voltages of these amplifiers are also V_c (assuming that the amplifiers are ideal). So the voltage across resistor R_a will be zero, which means that there will be a zero current through R_a and therefore no current through resistors R_b. This means that $V_1' = V_1$ and $V_2' = V_2$: the configuration behaves exactly like that of the original differential amplifier seen in Figure 12.5 as far as common mode signals are concerned.

Let us now consider a pure differential mode signal: $V_1 = -V_2 = \frac{1}{2} V_d$ or $V_1 - V_2 = V_d$. The voltage across resistor R_a is also equal to V_d, which means that a current through R_a will be equal to V_d/R_a. This current will flow through the resistors R_b producing a voltage across them that is equal to $V_d(R_b/R_a)$. The respective inputs of the original differential amplifier will then be $V_1' = V_1 + (R_b/R_a)V_d$ and $V_2' = V_2 - (R_b/R_a)V_d$, so its differential input will be

$$V_1' - V_2' = V_d\left(1 + 2R_b/R_a\right) \tag{12.7}$$

In conclusion, the differential gain of the configuration is increased by a factor of $1 + 2R_b/R_a$, while the common mode gain remains unaltered. The CMRR of the total configuration is thus increased by factor $1 + 2R_b/R_a$ as far as the original differential amplifier goes.

The differential mode gain can be adjusted by just a single resistor, R_b, which can be distinguished from the circuit shown in Figure 12.5 where two resistors have to be equally varied to change the gain and to maintain a reasonable CMRR.

12.2 Non-ideal operational amplifiers

An actual operational amplifier deviates from the ideal behavior in a number of ways. Such deviations limit applicability and require careful evaluation of the specifications before a proper choice can be made. It is important to know precisely what are the limitations of operational amplifiers and to know how their influence on the designed configuration can be gauged. First we shall discuss the main specifications of operational amplifiers. We shall then go on to study the effect they have on the properties of the circuits discussed in the preceding section while providing, where possible, solutions to the undesired effects.

12.2.1 The specifications of operational amplifiers

The specifications of the various types of operational amplifiers that are available are very divergent. Table 12.1 gives an overview of the main properties of three different types.

The specifications represent typical average values and are valid for temperatures of 25 °C. Minimum and maximum values are also often specified (see Appendix B.2.1).

What characterizes type I, given in Table 12.1, is its low price. Type II is noted for its excellent input characteristics (low offset voltage and low bias current) while type III is designed for high frequency applications.

The input components of type I are bipolar transistors while those of types II and III are JFETs. This is something that can also be deduced from the values of the input bias currents which correspond to the base current and gate currents of the transistors that are used. The favorable high frequency characteristics of type III are reflected in the high value of f_t and the slew rate.

Table 12.1. A selection of typical specifications for three types of operational amplifiers.

parameter	type I	type II	type III	Unit
V_{off}	1	0.5	0.5	mV
t.c. V_{off}	20	2 – 7	10	μV/K
I_{bias}	80	0.01	0.2	nA
I_{off}	20	0.002	0.02	nA
t.c. I_{off}	0.5	doubles per 10K		nA/K
input noise	4	25	20	nV/√Hz
		0.01		pA/√Hz
A_0	$2 \cdot 10^5$	$2 \cdot 10^5$	1500	-
R_i	$2 \cdot 10^6$	10^{12}	10^{12}	Ω
CMRR	90	80 – 100	100	dB
SVRR	96	80 –100	70	dB
f_t	1	1 – 4	300	MHz
slew-rate	0.5	3 – 15	400	V/μs

Explanation of the terms	
V_{off}	input offset voltage
I_{bias}	bias current
I_{off}	current offset, the difference between both bias currents
t.c.	temperature coefficient
A_0	DC gain
R_{i}	resistance between input terminals
SVRR	Supply Voltage Rejection Ratio
f_{t}	unity gain bandwidth
slew-rate	max. value of dV_o/dt

Table 12.2 provides an example of the maximum environmental conditions permitted with type I.

Table 12.2. The absolute maximal ratings for some of the parameters of type I given in Table 12.1.

power supply	±18 V
power dissipation	500 mW
input differential voltage	±30 V
output short-circuit duration	Indefinite
operating temperature range	0 … 75°C
storage temperature range	–65 … 150 °C
soldering temperature (60 s)	300 °C

12.2.2 *Input offset voltage*

One of the biggest possible limitations that can be placed upon an operational amplifier is the offset voltage, in particular when processing small DC signals. Both the input offset voltage V_{off} and the input bias currents I_{bias} contribute to the total offset of the circuit. This will be illustrated in Figure 12.7.

We model the input offset voltage and the bias currents as external sources, V_{off}, in series with one of the input terminals (it does not matter which of the two) and I_{bias} parallel to the input terminals. These sources account for all the errors taken into consideration which is why the operational amplifier itself is made error-free, the input currents are zero and the input voltage is zero. The input is usually connected to a low impedance voltage source. To calculate the errors the input terminal is therefore grounded.

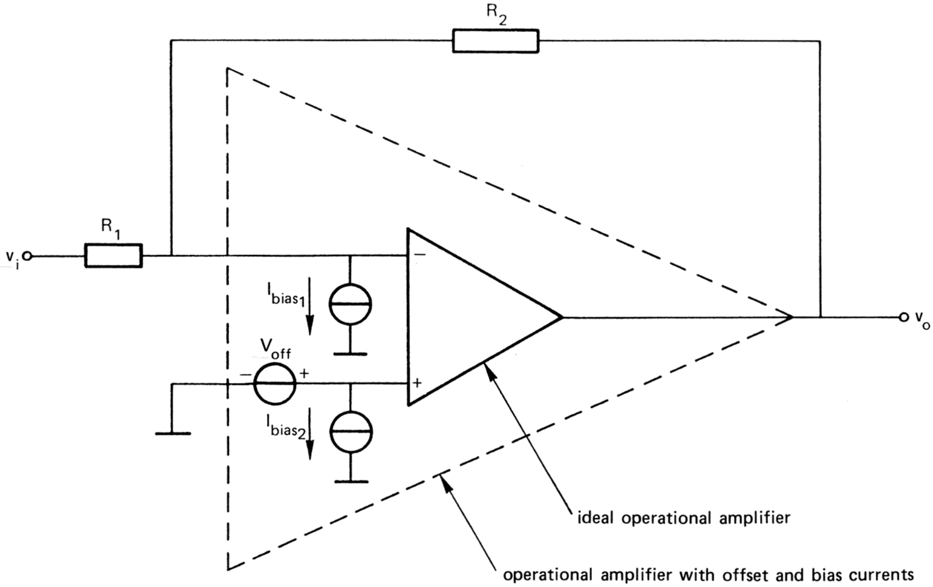

R2
R1
vi
vo
Ibias1
Voff
Ibias2
ideal operational amplifier
operational amplifier with offset and bias currents

Figure 12.7. The inverting amplifier of Figure 12.4b, including an offset voltage source and two bias current sources.

The calculation is made in two stages. First just the offset voltage contribution is calculated then just the bias current contribution is worked out. The resulting outputs are added up in accordance with the principle of superposition for linear systems and the output voltage is then divided by the voltage gain in order to find the equivalent total input error voltage.

First $I_{bias} = 0$, the output voltage due to V_{off} is $V_{off}\,(1 + R_2/R_1)$. Next $V_{off} = 0$ and the voltage on the inverting input is also zero. This means that there is no current through resistor R_1 (both terminals are at zero potential). The bias current I_{bias1} must go somewhere and the only path is through R_2. This leads to an output voltage that is equal to $I_{bias1}R_2$. The other bias current, I_{bias2}, flows directly to ground and does not contribute to the output voltage. The total output error voltage is:

$$V_{off,o} = \left|V_{off}\right|\left(1 + R_2/R_1\right) + \left|I_{bias1}\right|R_2 \qquad (12.8)$$

We do not know what the sign for V_{off} and I_{bias} is so we take the modulus to find the maximum worst case error signal. After dividing this by the voltage gain $-R_2/R_1$, we find that the maximum equivalent input error voltage is:

$$V_{off,i} = \left(1 + R_1/R_2\right)\left|V_{off}\right| + R_1\left|I_{bias1}\right| \tag{12.9}$$

Likewise, the error voltage of other arbitrary circuits can also be calculated.
The bias current of the operational amplifiers given in Figure 12.4 will flow entirely through the input voltage source connected to the circuit. That will create an extra error voltage of $I_{bias}.R_g$, where R_g is the source impedance. An amplifier with very low bias current is needed to measure the voltage from a source with a relatively high resistance so that the error voltage introduced is not too large.
The effect of both the offset and the bias current can be significantly reduced. The offset voltage can be reduced or compensated by means of an additional voltage added to the input. This compensation voltage can be taken from the power supply voltage (Figure 12.8a) by using a proper voltage divider circuit.

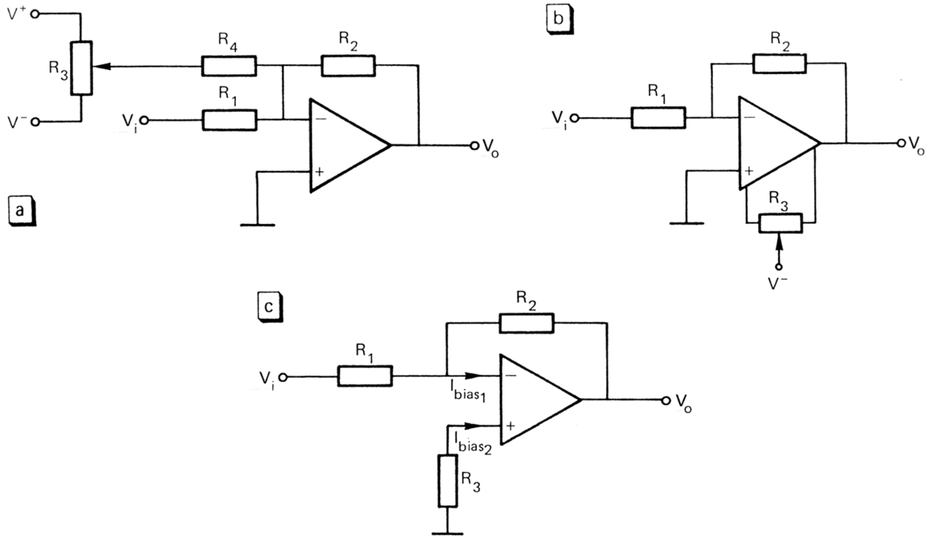


Figure 12.8. Various compensation methods: (a) with additional, adjustable voltage derived from the power supply, (b) with special operational amplifier connections, (c) bias current compensation.

Some operational amplifiers have special connections which may be used to realize compensation circuits more readily (Figure 12.8b). Bias current compensation is based on the principle that both bias currents are almost equal and have the same sign (they either flow to or from the amplifier). A resistor R_3 is connected in series with the non-inverting input (see Figure 12.8c). This creates extra offset voltage, $-I_{bias2}R_3$ at the non-inverting input point resulting in an extra output error voltage of $-I_{bias2}R_3$ $(1 + R_2/R_1)$. The contribution of $-I_{bias1}$ to the output is, as we have seen before, $I_{bias1}R_2$. If both bias currents are equal, the total output error voltage will be zero if R_3 satisfies the condition $R_2 = R_3(1 + R_2/R_1)$ or $R_3 = R_1//R_2$. Evidently the bias currents are not exactly equal, their difference is the bias offset current I_{off}. Under the same conditions for R_3, the output error voltage is equal to $(I_{bias1} - I_{bias2})R_2 = I_{off}R_2$, and the equivalent input error voltage is $I_{off}R_1$. As the offset current is usually small compared to the bias currents, the effect of the bias currents can largely be eliminated by simply adding resistor R_3.

These compensation methods are also applicable to most other operational amplifier circuits. After carefully compensating or adjusting the errors, the effect of the temperature coefficients for V_{off} and I_{off} will remain, which is why these parameters are specified as well.

Example 12.3

An inverting voltage amplifier in the configuration given in Figure 12.8c uses a 741 type operational amplifier (type I in Table 12.1). $R_1 = 10\ k\Omega$, $R_2 = 1\ M\Omega$ *and* $R_3 = R_1//R_2$. *At 20 °C the output voltage is adjusted to zero (at zero input voltage) using the compensation circuit seen in Figure 12.8b. We derive the maximum possible input error voltage in a temperature bracket ranging from 0 to 70 °C.*

The error input voltage is $V_{off,o} = 1.01|V_{off}| + 10^4|I_{off}|$. *At 20 °C these contributions cancel (due to adjustment). The maximum offset occurs at 70 °C:* $|V_{off}| = \Delta T \times t.c.(V_{off}) = 50 \times 20\ \mu V$; $|I_{off}| = \Delta T \times t.c.(I_{off}) = 50 \times 0.5\ nA$, *which is why at that temperature* $V_{off,i}$ *can be as high as 1.26 mV.*

12.2.3 Finite voltage gain

Up until now we have assumed that the operational amplifier has an infinite voltage gain. At low frequencies this assumption is acceptable but not in the case of higher frequencies. Figure 12.9 shows the amplitude transfer characteristic of a typical low-cost operational amplifier (type 741).

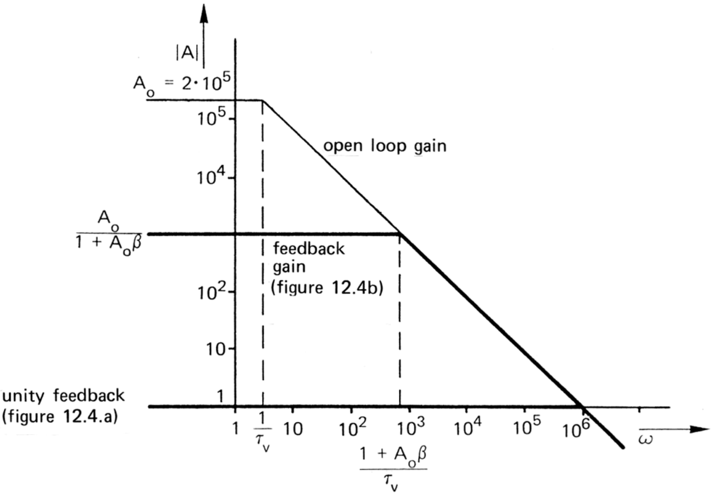


Figure 12.9. The amplitude transfer characteristic of an operational amplifier with and without feedback.

The characteristic has a remarkably low value of –3dB frequency. Is it still justifiable to assume that the gain is infinite? We will investigate this in the case of the non-inverting voltage amplifier shown in Figure 12.4b. The next equations apply to this configuration: $V_i = V^+$ (the voltage at non-inverting input); $V^- = V_o(R_1/(R_1 + R_2) = \beta V_o$ (where β is the fraction of the output that is fed back to the input); $V_o = A(V^+ - V^-)$. The elimination of V^+ and V^- will ultimately result in:

$$\frac{V_o}{V_i} = \frac{A}{1 + A\beta} \tag{12.10}$$

Only where $A\beta$ is large compared to 1, may we approximate the transfer

$$\frac{V_o}{V_i} = \frac{1}{\beta} = 1 + \frac{R_2}{R_1} \tag{12.11}$$

as found earlier. A less rough approximation would be:

$$\frac{V_o}{V_i} = \frac{A}{1 + A\beta} = \frac{1}{\beta}\frac{1}{1 + \frac{1}{A\beta}} \approx \frac{1}{\beta}\left(1 - \frac{1}{A\beta}\right) \tag{12.12}$$

The relative deviation from the ideal value is about $1/A\beta$. The term $A\beta$ is often encountered in control system calculations. It is called the open loop system gain and it is the amplifier transfer and the feedback network put together. The higher the loop gain, the less the transfer will be affected by the amplifier itself. This is the main reason for striving for the highest possible operational amplifier voltage gain. With the buffer amplifier given in Figure 12.4a, $\beta = 1$ the actual transfer deviates only a fraction $1/A$ from 1.

As can be seen in Figure 12.9, A decreases at increasing frequency. Apparently the amplifier behaves like a first order low-pass filter with a complex transfer function

$$A(\omega) = \frac{A_0}{1 + j\omega\tau_v} \tag{12.13}$$

Substituting this expression for the transfer function of the non-inverting amplifier configuration will lead to:

$$\frac{V_o}{V_i} = \frac{A(\omega)}{1 + A(\omega)\beta} = \frac{A_0}{1 + j\omega\tau_v + A_0\beta} = \frac{A_0}{1 + A_0\beta}\frac{1}{1 + \frac{j\omega\tau_v}{1 + A_0\beta}} \tag{12.14}$$

This transfer has a first-order low-pass characteristic as well: the -3dB frequency occurs at $\omega = (1 + A_0\beta)/\tau_v$. The bandwidth is factor $1 + A_0\beta$ larger than that of the open amplifier. The low-frequency transfer is $A_0/(1 + A_0\beta)$ and it is factor $1 + A_0\beta$ less than that of the open amplifier. Obviously the bandwidth and gain product remains the same, irrespective of the feedback. This can also be seen in Figure 12.9. At unity feedback, $\beta = 1$, so $V_o/V_i = A_0/(1 + A_0) \approx 1$. The bandwidth is $(1 + A_0)/\tau_v \approx A_0/\tau_v$. This is the unity gain bandwidth which is denoted as f_t (in Hz). From the unity gain bandwidth it is immediately clear what is the bandwidth of a circuit with arbitrary gain.

> ***Example 12.4***
> *An operational amplifier has a unity gain bandwidth of f_t = 2 MHz. The bandwidth of an amplifier with a gain of 100 is 20 kHz. At a gain of 1000 the bandwidth is only 2 kHz.*

Only amplifiers that behave like first-order low-pass filters have a constant gain-bandwidth product (GB-product). Many amplifiers, in particular those used for high frequency applications, have a second-order or even higher-order transfer function. Such amplifiers are not always stable at arbitrary feedback. However, they do have some extra connections so that compensating networks, like for instance a small capacitor, can be added to guarantee stability at the chosen gain factor. This is called external frequency compensation. The manufacturer provides the necessary instructions in the relevant specification sheets.

As a point of interest, first-order amplifiers were also originally of a higher order. The first-order low −3dB frequency characteristic derives from internal frequency stabilization provided to guarantee arbitrary gain stability.

SUMMARY

Amplifier circuits with ideal operational amplifiers

- An ideal operational amplifier has an infinite voltage gain, CMRR and input impedance. The voltage offset, bias currents and output impedance are zero.
- At proper feedback, the voltage between the two input terminals of an operational amplifier is zero.
- Operational amplifiers are used to realize various signal operations. Those operations are fixed by external components and the properties are virtually independent of the properties of the operational amplifier itself.
- The voltage transfer of the inverting voltage amplifier seen in Figure 12.3a is $-R_2/R_1$ while that of the non-inverting amplifier given in Figure 12.4b is $1 + R_2/R_1$.
- The inverting voltage amplifier has fairly low input resistance, R_1, the input resistance of the non-inverting amplifier is very high.
- The differential gain of the differential amplifier in Figure 12.5 is $-R_2/R_1$, under the condition $R_1 = R_3$ and $R_2 = R_4$.
- The CMRR and the input resistance of the differential amplifier in Figure 12.5 can be substantially increased by adding two operational amplifiers arranged in the way shown in Figure 12.6.

Non-ideal operational amplifiers

- The output offset voltage of the inverting or non-inverting voltage amplifier comprises the terms $(1 + R_2/R_1)V_{off}$ and $I_{bias}R_2$. The contribution of V_{off} can be reduced by employing the compensation methods shown in Figures 12.8a and b. The contribution of I_{bias} is reduced to $I_{off}R_2$ by adding the compensation resistor R_3 in accordance with Figure 12.8c. These compensation methods are also applicable to other amplifier configurations.
- The equivalent input offset voltage or error voltage is the output error voltage divided by the voltage transfer.
- An operational amplifier with a first-order amplitude transfer characteristic has a constant gain-bandwidth product which is specified as the unity gain bandwidth f_t.

EXERCISES

Amplifier circuits with ideal operational amplifiers

12.1 What is meant by a virtual ground? The minus-terminal of an inverting amplifier is virtually grounded, why? The minus-terminal of a non-inverted amplifier is not virtually grounded. Explain this.

12.2 All operational amplifiers in Figures a-f (next page) can be viewed as ideal. Find the output voltage V_o.

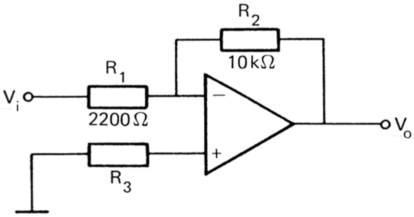
12.3 Design an amplifier in accordance with Figure 12.3a which has a voltage gain of −50 and a minimum input resistance of 5 kΩ. Only take resistance values from the E12 series.

12.4 Design a circuit that has just one operational amplifier and which is able to add three voltages so that $V_o = -10V_1 - 5V_2 + 2V_3$.

12.5 Deduce the transfer for the common mode and differential mode signals for the circuit given in Figure 12.6 where the two resistors R_b are not equal. For a high CMRR is it necessary for them to be equal?

Non-ideal operational amplifiers

12.6 The specifications for the operational amplifier given in the circuit below are: V_{off} = 0.4 mV; I_{bias} = 10 nA; I_{off} = 1 nA; f_t = 5 MHz. Calculate V_o due to V_{off} only.

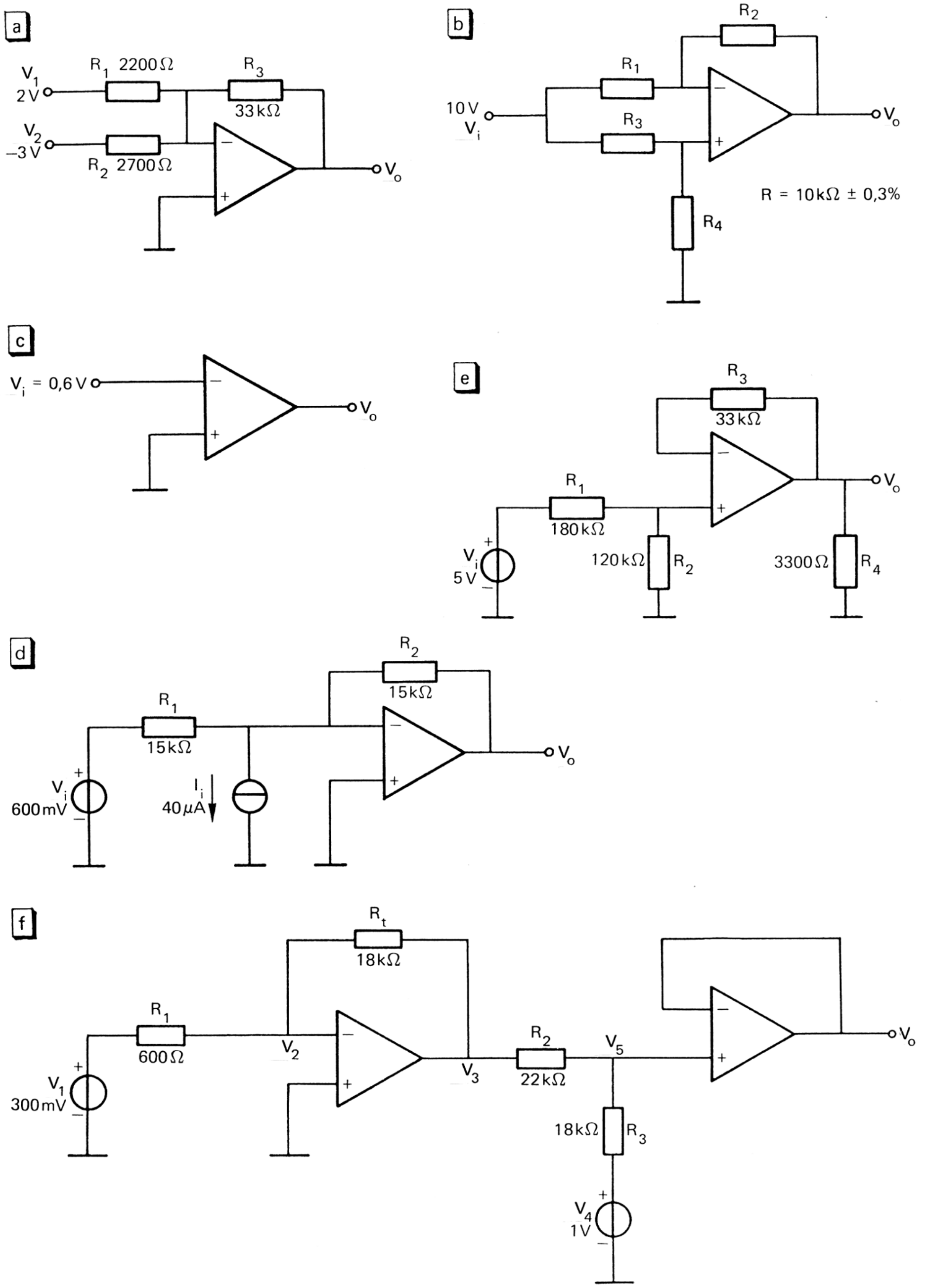


12.7 Calculate the V_o in the circuit used in exercise 12.6 which is due only to I_{bias}. Find the proper value of R_3 for optimal compensation.

12.8 Find the input resistance in the circuit shown in exercise 12.6. Try also to determine the output resistance.

12.9 Find the bandwidth of the same circuit.

12.10 Design a circuit in accordance with the details given in Figure 12.8a so that the voltage gain is −30 and the input resistance is at least 10 kΩ. The output offset voltage should be adjustable in the range of −1.2 to +1.2 V. The power supply voltages are +15 and −15 V.


a
V1 2V
R1 2200Ω
R3 33kΩ
V2 −3V
R2 2700Ω
Vo
b
R2
R1
R3
10V Vi
Vo
R4
R = 10kΩ ± 0,3%
c
Vi = 0,6V
Vo
e
R3 33kΩ
R1 180kΩ
Vi 5V
120kΩ R2
3300Ω R4
Vo
d
R2 15kΩ
R1 15kΩ
Vi 600mV
Ii 40μA
Vo
f
Rt 18kΩ
R1 600Ω
V1 300mV
V2
V3
R2 22kΩ
V5
18kΩ R3
V4 1V
Vo

13 Frequency selective transfer functions with operational amplifiers

In Chapter 8 several circuits for frequency selective signal processing with passive components were discussed. Certain of the disadvantages of such circuits, like the unfavorable input and output impedances or the obligatory use of inductors, can be overcome by using operational amplifiers.

The first part of this chapter describes circuits for signal processing functions in the time domain, such as the integrator and the differentiator. The second part deals with filter circuits that have a high selectivity.

13.1 Circuits for time domain operations

In this section the basic configurations for circuits are the inverting and the non-inverting amplifiers discussed in the previous chapter (Figure 13.1). The complex transfer functions of these circuits are expressed as $H = -Z_2/Z_1$ and $H = (Z_1 + Z_2)/Z_1$, respectively.

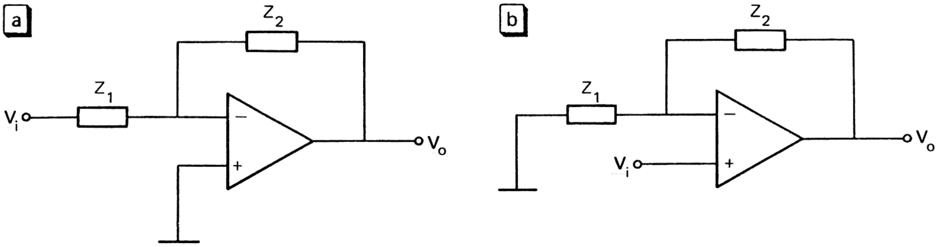


Figure 13.1. The basic configurations of the circuits in this chapter, (a) with inverting transfer and (b) with non-inverting transfer.

13.1.1 The integrator

The circuit in Figure 13.1a acts as an electronic integrator when Z_1 is a resistance and Z_2 a capacitance (Figure 13.2).

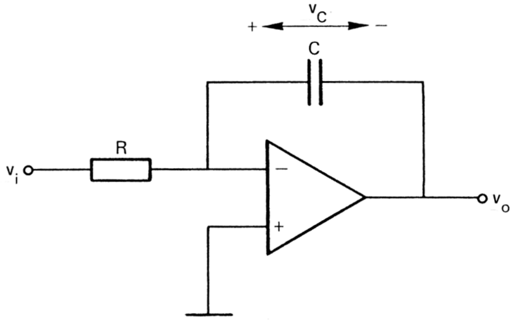


Figure 13.2. Basic electronic integrator.

The transfer function of this circuit is $H = -1/j\omega RC = -1/j\omega\tau$. Its modulus equals $1/\omega\tau$, which is inversely proportional to the frequency: the modulus thus decreases by 6 dB/octave over the full frequency range. If one disregards the minus sign in the transfer, its argument has a constant value of $-\pi/2$. A sinusoidal input signal $v_i = \hat{v}\sin\omega t$ gives an output signal equal to $v_o = -(1/\omega\tau)\hat{v}\sin(\omega t - \pi/2) = -(1/\omega\tau)\hat{v}\cos\omega t$, which is the integrated value of v_i.

This circuit acts as an integrator for other signals as well and can be shown in the way described below. Because of the virtual ground of the inverting input terminal, the input signal v_i is converted into a current v_i/R, which flows through the capacitance C. The output of the amplifier is $v_o = -v_c$. The current through a capacitor is $i = C(dv_c/dt)$. Since $i = v_i/R$, it follows that $C(dv_c/dt) = v_i/R$, so the output voltage satisfies the equation

$$v_o = -\int \frac{v_i}{RC}dt = -\frac{1}{\tau}\int v_i dt \tag{13.1}$$

This only holds for an ideal operational amplifier. Imperfections in the amplifier cause deviations from the ideal integrator behavior. The influences of the bias current I_{bias} and the offset voltage V_{off} can be deduced from Figure 13.3. First one must suppose that $I_{bias} = 0$. The current through R equals V_{off}/R as the non-inverting input is virtually grounded. This current flows through the capacitance C. After that in order to find the effect of I_{bias}, the offset V_{off} is presumed to be zero. If that is indeed the case I_{bias} will flow directly into C because the voltage across R is zero. At zero input the total current through C is therefore $I_{bias} + V_{off}/R$, which will result in an output voltage equal to

$$v_{off,o} = -\frac{1}{C}\int\left(I_{bias} + \frac{V_{off}}{R}\right)dt + V_{off} \tag{13.2}$$

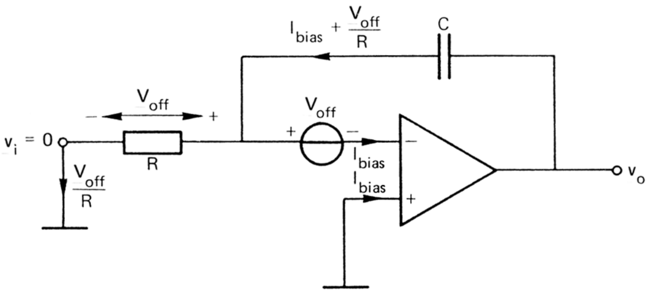

$I_{bias} + \frac{V_{off}}{R}$
C
V_{off}
−
+
$v_i = 0$
R
$\frac{V_{off}}{R}$
V_{off}
+
−
I_{bias}
I_{bias}
−
+
v_o

Figure 13.3. To determine the effect of the offset voltage V_{off} and the bias current I_{bias} the input signal is set at zero.

Most of the polarities of I_{bias} and V_{off} are unknown so the worst case output offset is

$$v_{off,o} = -\frac{1}{C}\int\left(\left|I_{bias}\right| + \frac{\left|V_{off}\right|}{R}\right)dt + \left|V_{off}\right| \tag{13.3}$$

Example 13.1
An integrator is designed in the way given in Figure 13.2, with R = 10 kΩ and C = 0.01 μF. The operational amplifier that is used in this circuit has an offset voltage that is less than 1 mV and a bias current that is less than 100 nA. On the basis of the last equation the output will increase by 20 volts each second so within 1 second after switching on, the amplifier will be in saturation.

To prevent the circuit from becoming saturated as a result of offset voltage and bias current one must select an amplifier with low values of V_{off} and I_{bias}. Besides that the following measures can be taken (see Figure 13.4):

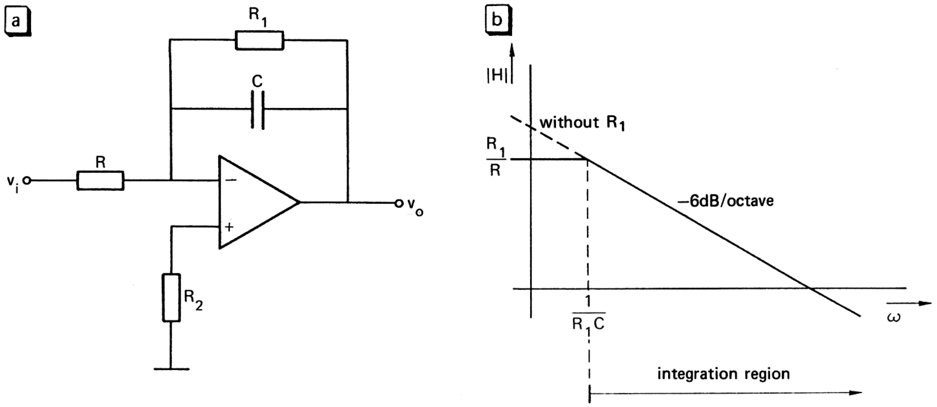

a
R_1
C
R
v_i
−
+
v_o
R_2
b
|H|
without R_1
$\frac{R_1}{R}$
−6dB/octave
$\frac{1}{R_1 C}$
ω
integration region

Figure 13.4. (a) The resistor R_1 in this circuit reduces the effect of the offset voltage and R_2 compensates for the bias current, (b) the transfer characteristic shows that the circuit only acts as an integrator for frequencies much higher than $1/(2\pi R_1)$ Hz.

- An additional resistor R_2 in series with the non-inverting input of the operational amplifier can be implemented. This reduces the effect of the bias current (Section 12.2.2).
- Offset compensation, similar to that of the inverting amplifier (see, for instance, Figures 12.8a and b), can be introduced.
- An additional resistor R_1 in parallel to C can be used. The transfer will then become $-(R_1/R)/(1 + j\omega R_1C)$. Figure 13.4b shows the transfer characteristic. At low frequencies, the transfer is limited to R_1/R; the integrator is then said to be tamed. A disadvantage of this method is its limited integrating range. Only signals with frequencies much higher than $1/(2\pi R_1C)$ Hz are integrated.

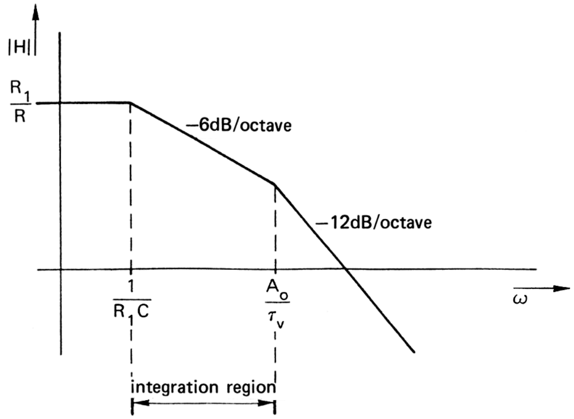


Figure 13.5. The frequency characteristic of the integrator in Figure 13.4a in the case of a band-limited operational amplifier. The working range of the integrator lies within the two –3dB frequencies.

The finite gain of the operational amplifier limits the usable range of the integrator at low frequencies and the finite bandwidth sets an upper limit for the integration range. The integration range is found as follows. The frequency-dependent gain of the amplifier is (see Section 12.2.3):

$$A(j\omega) = \frac{A_0}{1 + j\omega\tau_v} \tag{13.4}$$

with A_0 as the DC gain or open loop gain and τ_v as the first-order time constant of the amplifier. The transfer function of the integrator in Figure 13.4a can be deduced from the following equations:

$$V_o = \frac{A_0}{1 + j\omega\tau_v}\left(V^+ - V^-\right) \tag{13.5}$$

$$\frac{V_i - V^-}{R} = \frac{V^- - V_o}{Z_1} \tag{13.6}$$

where $Z_1 = R_1/(1 + j\omega R_1C)$, and V^- and V^+ the respective voltages at the inverting and non-inverting inputs of the amplifier. V^+ is zero because the current through R_2 is very small. If one eliminates V^- from these two equations under the conditions $A_0 >> 1$, $A_0R >> R_1$, $R_1 >> R$ and $R_1C >> \tau_v/A_0$, the transfer of the integrator circuit may approximate:

$$\frac{V_o}{V_i} = -\frac{R_1}{R} \cdot \frac{1}{1 + j\omega R_1 C} \cdot \frac{1}{1 + j\omega\tau_v/A_0} \tag{13.7}$$

The frequency characteristic is depicted in Figure 13.5. Apparently the circuit integrates signals for which $1/R_1C << \omega << A_0/\tau_v$.

Example 13.2
For the circuit given in Figure 13.4, the components have the following values: R_1 = 1 MΩ, C = 0.1 μF, R = 1 kΩ, $A_0 = 10^5$ and τ_v = 0.1 s. In this design, the cut-off frequencies are 10 and 10^6 rad/s. This means that there is an integration range from roughly 15 Hz to 150 kHz.

13.1.2 Differentiator

By interchanging the positions of the resistance and the capacitance in Figure 13.2, the circuit turns into the differentiator shown in Figure 13.6.

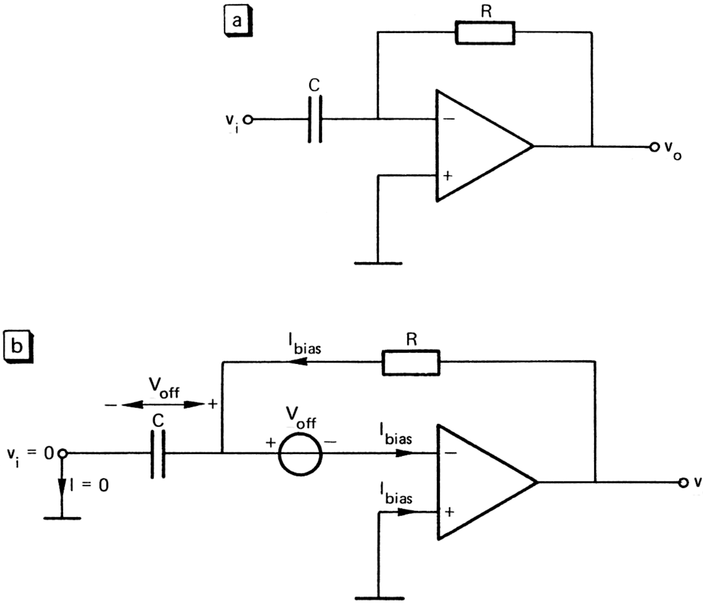


Figure 13.6. (a) Basic electronic differentiator, (b) to determine the effect of the offset voltage V_{off} and the bias current I_{bias} the input signal is set at zero.

The input voltage v_i is converted into a current $i = C(dv_i/dt)$ flowing through capacitor C. This current flows into the feedback resistor R and results in an output voltage equal to $v_o = -iR = -RC(dv_i/dt)$.

The transfer in the frequency domain is $H = -Z_2/Z_1 = -j\omega\tau$. The modulus of the transfer is directly proportional to the frequency (a rise of 6 dB per octave) and the argument (disregarding the minus sign) is $\pi/2$ over the whole frequency range.

Just as in the case of the integrator, the differentiator also suffers from the non-ideal characteristics of the operational amplifier, thus limiting its useful range of operation. The bias current causes a constant output voltage equal to $-I_{bias}R$. The offset voltage appears unchanged at the output, because there is no current through R (at zero bias current), see Figure 13.6b. If the polarity of V_{off} and the direction of I_{bias}, are not known the output offset voltage will be less than $|V_{off}| + |I_{bias}|R$, which is a small and constant value.

A serious disadvantage of this circuit, however, is the high gain for signals with high frequency components. Wide band thermal noise and rapidly changing interference signals are highly amplified by the differentiator. Furthermore, steep input signals (like those from square or pulse-shaped voltages) can saturate the differentiator. Finally, the frequency-dependent transfer of the operational amplifier may give rise to a pronounced peak in the frequency characteristic of the circuit or even to instability.

For all these reasons, a differentiator of the type given in Figure 13.6 is not recommended. The circuit shown in Figure 13.7 behaves better. There the range of the differentiator is restricted to low frequencies only. By adding R_1, the high-frequency gain is limited to a certain maximum. If the frequency dependence of the operational amplifier is not taken into account, the transfer function of the circuit given in Figure 13.7 will satisfy:

$$H(j\omega) = -\frac{Z_2}{Z_1} = -\frac{R}{R_1 + 1/j\omega C} = -\frac{j\omega RC}{1 + j\omega R_1 C} = -\frac{j\omega\tau}{1 + j\omega\tau_1} \qquad (13.8)$$

with $\tau_1 = R_1C$. The frequency characteristic of this transfer is shown in Figure 13.7b.

> ***Example 13.3***
>
> *The aim is to design a differentiator for frequencies up to 100 Hz (about 600 rad/s); the gain may not be more than 10. These requirements will be met if, for instance,* $\tau_1 = R_1C = 10^{-3}$ *s. The* R/R_1 *ratio should be 10. For a stable transfer it is favorable to have a low gain at high frequencies. A suitable design is* $R_1 = 10\ \text{k}\Omega$, $R = 100\ \text{k}\Omega$ *and* $C = 0.1\ \mu\text{F}$.

13.1.3 Circuits with PD, PI and PID characteristics

Control systems often require a specific transfer characteristic. Such a characteristic may contain a proportional region (P, a frequency-independent transfer), a differentiating region (D) and an integrating region (I), or a combination of the three (denoted as PI, PD and PID). These names reflect the type of transfer in the time domain of the controller that generates the control signals. All these characteristics may be realized with the configuration given in Figure 13.1a and they are briefly discussed below (Figure 13.8).

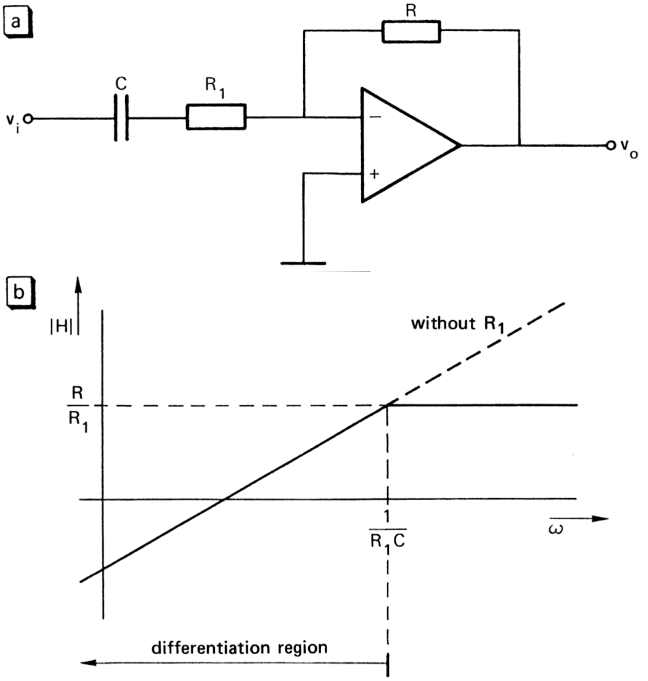


Figure 13.7. (a) To reduce noise and high-frequency interference, resistor R_1 is added to the circuit shown in Figure 13.6, (b) the frequency characteristic of the differentiator shows that its range is restricted to frequencies below $1/(2\pi R_1 C)$.

The PI-characteristic (Figure 13.8a)

$Z_1 = R_1$ and $Z_2 = R_2 + 1/j\omega C_2$. The transfer becomes $H = -(R_2/R_1 + 1/j\omega R_1 C_2)$. If the integrator is to be prevented from saturating, the gain for AC signals must be restricted (see the upper dotted line in Figure 13.8a).

The PD-characteristic (Figure 13.8b)

Z_1 consists of a resistor R_1 in parallel to capacitor C_1; $Z_2 = R_2$. The transfer equals $H = -(R_2/R_1 + j\omega R_2 C_1)$. As discussed in the paragraph on the differentiator, measures have to be taken to guarantee the stability of the system (see dotted line).

The PID-characteristic (Figure 13.8c)

$Z_1 = R_1/(1 + j\omega R_1 C_1)$; $Z_2 = R_2 + 1/j\omega C_2$. The transfer function satisfies $H = -(C_1/C_2 + R_2/R_1 + 1/j\omega R_1 C_2 + j\omega R_2 C_1)$.
To restrict the working range of the integrator to high frequencies only, a resistor R_p is adjusted in parallel to Z_2. The working range of the differentiator is restricted to low frequencies by a resistor R_s in series with C_1. The circuit is depicted in Figure 13.8d.

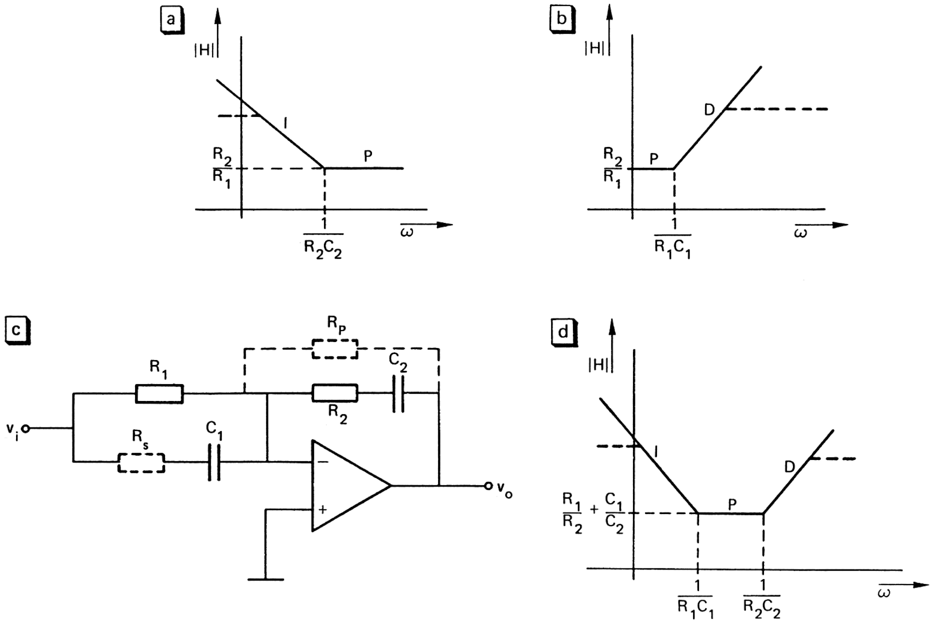


Figure 13.8. Transfer characteristics of (a) a PI-circuit, (b) a PD-circuit, (c) a PID circuit, (d) a circuit for a PID-characteristic.

The low output impedance of the circuits discussed above allows several circuits to cascade without there being significant loading effects . The (complex) transfer function of such a serial system is simply the product of the individual transfer functions. For other filter characteristics, special filter characteristics and the respective pulse and step responses, the reader is referred to general text books on this subject.

13.2 Circuits with high frequency selectivity

This section shows how to effect a high selectivity transfer function without using inductors and by applying active components (operational amplifiers). We shall discuss second-order band-pass filters with a very small bandwidth and low-pass filters of a higher order.

13.2.1 Resonance filters

Passive filters consisting of inductors and capacitors can have a high selectivity (or a very small bandwidth) in conjunction with the effects of resonance. To realize resonance effects inductorless band-pass filters that have a high selectivity require the use of active components, such as operational amplifiers.

The general transfer function of a second order network can be presented as:

$$H(j\omega)=\frac{a_0+a_1(j\omega)+a_2(j\omega)^2}{b_0+b_1(j\omega)+b_2(j\omega)^2} \tag{13.9}$$

The coefficients a_0, a_1 and a_2 in this expression are not all equal to zero. If the numerator contains only the factor $a_1 j\omega$ the filter will be of the band-pass type, because the modulus of the transfer approaches zero for $\omega \rightarrow 0$ as well as for $\omega \rightarrow \infty$. If $a_2 = 0$ (and possibly also $a_1 = 0$), $|H| = a_0/b_0$ for $\omega = 0$, whereas for $\omega \rightarrow \infty$ the transfer will be near to zero. Such a filter can therefore be said to have a low-pass character. Similarly, it can be proven that the transfer has a high-pass character if $a_0 = 0$.

Band-pass type filters will be discussed, so $a_0 = a_2 = 0$. The denominator of $H(j\omega)$ can be rewritten as $1 + 2j\omega z/\omega_0 - \omega^2/\omega_0^2$ (see Section 6.1.2) or as $1 + j\omega/Q\omega_0 - \omega^2/\omega_0^{2.}$ In this section the latter expression will be used. If the numerator is written as $H_0 j\omega/Q\omega_0$ then the transfer function of the second-order band-pass filter becomes

$$H(j\omega) = H_0 \frac{j\omega/Q\omega_0}{1 + j\omega/Q\omega_0 - \omega^2/\omega_0^2} \qquad (13.10)$$

The transfer is fixed with three parameters: H_0, ω_0 and Q. H_0 is the transfer at $\omega = \omega_0$. This is equal to the maximum transfer. That is why ω_0 is called the resonance frequency. (In Section 6.1.2 this maximum transfer has been calculated for a second-order function with $a_0 = 1$ and $a_1 = a_2 = 0$. In such cases the maximum transfer occurs at slightly lower frequencies). The bandwidth of the filter (the frequency span between the two −3 dB points) can be worked out by calculating the frequencies $\omega \pm \Delta\omega$ for which $|H| = \frac{1}{2}H_0\sqrt{2}$. This bandwidth appears to be equal to $B = 2\Delta\omega = \omega_0/Q$, assuming that $\Delta\omega << \omega_0$ (see Figure 13.9). The parameter Q increases as the bandwidth decreases which is why Q is called the filter's quality factor (see also Section 6.1.2). With only passive components (resistors and capacitors), the maximum obtainable value for Q is 0.5.

The number of inductorless band-pass filters one can possibly have is considerable. We will restrict ourselves to just two classes, viz. filters created with one operational amplifier that have positive and negative feedback.

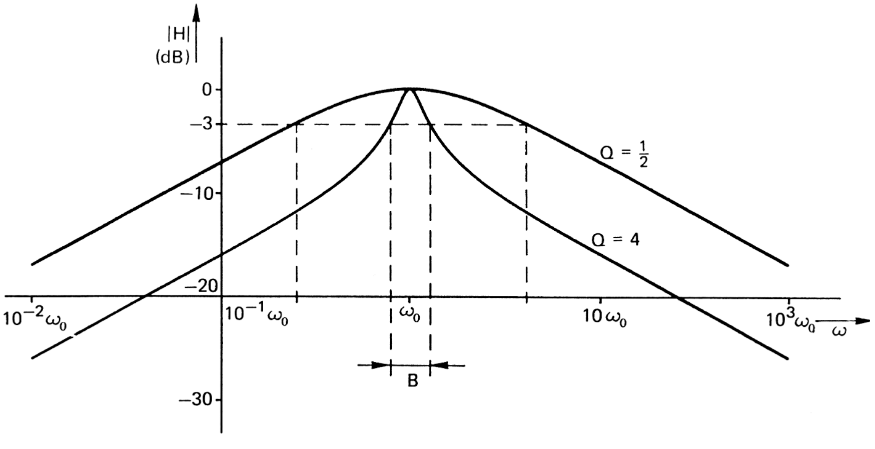


Figure 13.9. The frequency characteristic of a second-order band-pass filter with transfer $H(j\omega) = j\omega/\omega_0\, Q/(1 + j\omega/j\omega_0\, Q - (\omega/\omega_0)^2)$, for different values of Q.

13.2.1.1 Band-pass filters with frequency-selective positive feedback

Figure 13.10 shows the control diagram of an amplifier that has positive feedback. The feedback network consists of passive components and has a transfer equal to $\beta(\omega)$. The voltage transfer V_o/V_i can be calculated using the equations $V_o = K(V_i + V_t)$ and $V_t = \beta(\omega)V_o$, to yield

$$\frac{V_o}{V_i} = \frac{K}{1 - K\beta(\omega)} \tag{13.11}$$

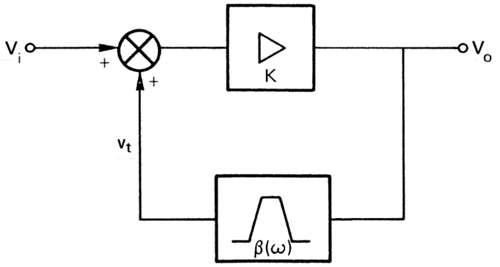


Figure 13.10. Basic operating of a system with positive feedback. The transfer function has a band-pass characteristic because the feedback network is a band-pass type filter.

If the feedback network has a band-pass characteristic (e.g. one of the networks from Section 8.1.3), then $H(\omega)$ also has a band-pass characteristic. The properties of such a band-pass filter are clarified in the next example given in Figure 13.11.

The K gain, the summation of the input signal and the feedback signal are realized in a single operational amplifier. The passive band-pass filter consists of the components R_1, R_2, C_1 and C_2. Assuming that the properties of the operational amplifier are ideal, the following equations will hold:

$$\frac{V_i - V^+}{Z_i} = \frac{V^+ - V_o}{Z_2} \tag{13.12}$$

$$V^- = \frac{R_3}{R_4 + R_3} V_o \tag{13.13}$$

$$V^- = V^+ \tag{13.14}$$

with $Z_1 = R_1 // j\omega C_1$ and $Z_2 = R_2 + 1/j\omega C_2$; V^- and V^+ being the respective voltages at the inverting and non-inverting operational amplifier inputs. When V^- and V^+ are eliminated from these equations, the transfer function is found to be equal to

$$H(j\omega) = \frac{V_o}{V_i} = \frac{R_4 + R_3}{R_3} \frac{(1 + j\omega\tau)^2}{1 + j\omega\tau(2 - R_4/R_3) - \omega^2\tau^2} \tag{13.15}$$

where it is assumed that $R_1 = R_2 = R$ and $C_1 = C_2 = C$. From the denominator it appears that $\omega_0 = 1/\tau$ and $Q = R_3/(2R_3 - R_4)$. The transfer at resonance is $H_0 = 2(R_4 + R_3)/(2R_3 - R_4)$. The sensitivity in Q to varying resistance values is high, in particular when the difference between $2R_3$ and R_4 is small. The same holds for H_0, the transfer at resonance point. The system is unstable for $R_4 \geq 2R_3$.

The resonance frequency and the quality factor of the filter can be varied independently: Q only depends on R_3 and R_4, whereas with R and C the resonance frequency can be tuned. A precondition for this independence is the equality of both R resistors and both C capacitances.

The filter in Figure 13.11 is called a Wien filter, because it is derived from a particular type of measuring bridge, the Wien bridge. Many other types of filters can be designed along the lines of the general principle demonstrated in Figure 13.10. A common property of these filters is the high sensitivity for varying parameters at high Q. That is why it is difficult to realize stable filters with a high quality factor. An advantage of these types of filters is the relatively low value of K: for the Wien filter the value is 3 at the very most. $K = 3$ corresponds to an infinite Q-factor. This low value allows such filters to be designed for relatively high frequencies (the gain-bandwidth product of an operational amplifier is constant and so a low gain corresponds to a high bandwidth).

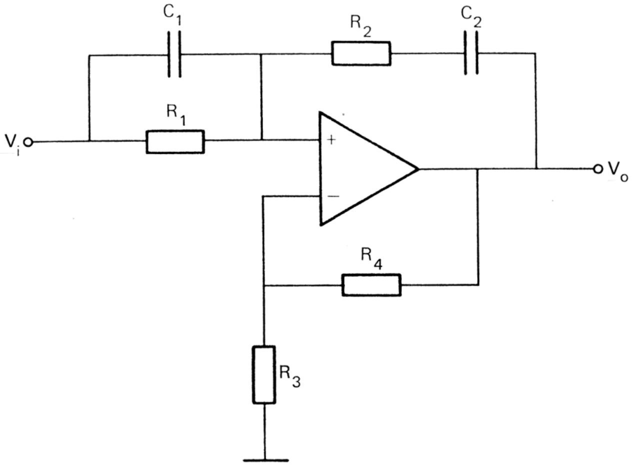


Figure 13.11. An example of a band-pass filter, made up of an operational amplifier with positive feedback.

13.2.1.2 Band-pass filters with frequency-selective negative feedback

A high quality factor can also be achieved with negative feedback. If that is to happen, the passive feedback network should have a notch characteristic. Figure. 13.12 shows the control diagram of such a system. Again, the operation will be explained by means of exemplification, see therefore Figure 13.13.

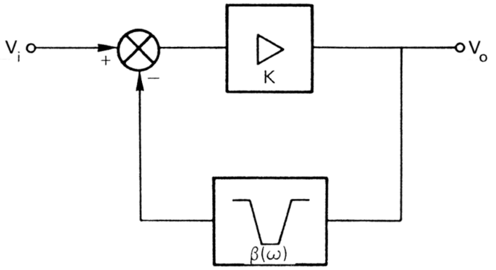


Figure 13.12. The basic principle of a filter configuration with negative feedback. The transfer function has a band-pass characteristic because the feedback network has a notch characteristic.

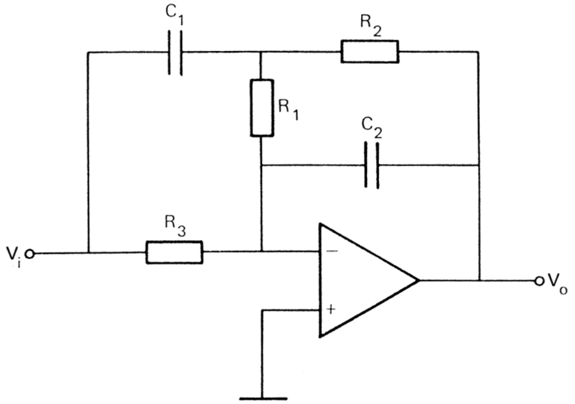


Figure 13.13. An example of a band-pass filter with an operational amplifier and frequency-selective negative feedback.

The notch filter in this circuit is hardly recognizable as such because the amplifier and filter functions are combined with a small number of passive components. The negative feedback is easily recognized: there is just one feedback path from the output to the inverting input of the operational amplifier. The transfer function of this filter appears to be

$$H(j\omega) = \frac{V_o}{V_i} = -\frac{1}{R_3} \cdot \frac{R_1 + R_2 + j\omega(R_1 + R_3)R_2C_1}{1 + j\omega(R_1 + R_2)C_2 - \omega^2 R_1 R_2 C_1 C_2} \qquad (13.16)$$

The resonance frequency is $\omega_0 = 1/\sqrt{(R_1R_2C_1C_2)}$, the quality factor is $Q = \sqrt{(R_1R_2C_1/C_2)}/(R_1 + R_2)$. The maximum value of Q is _$\sqrt{(C_1/C_2)}$, for $R_1 = R_2$. The quality factor is determined by the ratio of two capacitance values and is, therefore, very stable. A disadvantage of this filter is that the parameters ω_0 and Q cannot be varied independently, as can be seen from the expressions for ω_0 and Q .

An important property of the band-pass filter with negative feedback is its unconditional stability, also at high Q. On the other hand, a high Q requires a high K gain. When an operational amplifier is used, this means that the value will be restricted to ω_0, because the bandwidth of the amplifier with feedback decreases as the gain increases.

13.2.2 Active Butterworth filters

In Section 8.2, filter types with different approximations to the ideal behavior were examined. When using operational amplifiers these filters can be designed without inductors. The discussion in this section will be restricted to Butterworth filters of the second and third order.

A second-order low-pass filter is depicted in Figure 13.14. This is the so-called Sallen-and-Key filter. The conditions for a Butterworth characteristic can be derived from the transfer function. The transfer is described as:

$$H(j\omega)=\frac{V_o}{V_i}=\frac{1}{1+j\omega(R_1+R_2)C_2-\omega^2R_1R_2C_1C_2} \tag{13.17}$$

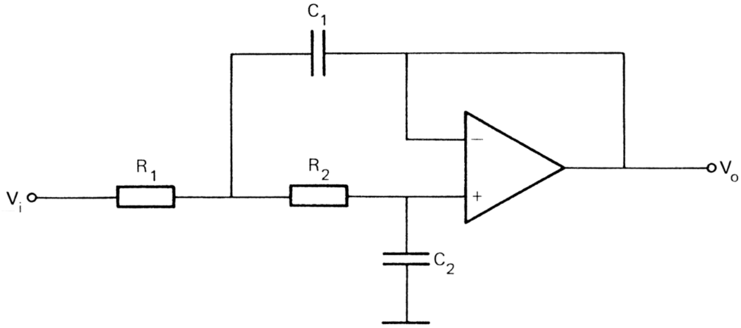


Figure 13.14. An example of an active second-order low-pass Butterworth filter.

The modulus of $H(j\omega)$ is

$$|H(j\omega)|=\frac{1}{\sqrt{\left(1-\omega^2R_1R_2C_1C_2\right)^2+\omega^2(R_1+R_2)^2C_2^2}} \tag{13.18}$$

It is a Butterworth characteristic (see Section 8.2.2) if the factor with ω^2 equals zero: $(R_1+R_2)^2C_2{}^2=2R_1R_2C_1C_2$, or $C_1/C_2=(R_1+R_2)^2/2R_1R_2$. To simplify the design, $R_1=R_2=R$, so $C_1=2C_2$.

The circuit can be extended to a third-order filter by adding a first-order low-pass section (Figure 13.15). At the output, an additional amplifier stage is connected, to achieve low output impedance and to make it possible to choose an arbitrary gain. The transfer function is calculated to determine the Butterworth condition of this filter. To simplify the calculation we directly propose $R_1=R_2=R_3=R$. The modulus of the transfer function then becomes:

$$|H(j\omega)| = \frac{1 + R_5/R_4}{\sqrt{\left\{\left(1 - \omega^2 R^2 C_1 C_2\right)^2 + 4\omega^2 R^2 C_2^2\right\}\left\{1 + \omega^2 R^2 C_3^2\right\}}} \tag{13.19}$$

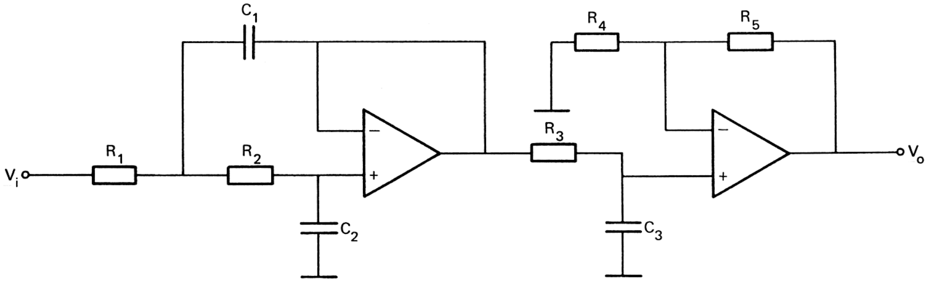


Figure 13.15. An example of an active, third-order low-pass filter with Butterworth characteristics that consists of second-order and first-order filters in cascade.

This expression has a Butterworth characteristic if the factors with ω^2 as well as with ω^4 are zero. This results in:

$$-2C_1C_2 + 4C_2^2 + C_3^2 = 0 \tag{13.20}$$

and

$$C_1^2C_2^2 - 2C_1C_2C_3^2 + 4C_2^2C_3^2 = 0 \tag{13.21}$$

Both equations are satisfied for $C_1 = 4C_2$ and $C_3 = 2C_2$.
It will be clear from the forgoing that the deriving of Butterworth conditions for filters of an even higher order becomes very time-consuming. Furthermore, the equations can no longer be solved analytically so numerical methods are required. In publications on this subject much information can be found on the design of higher order filters and other types of filters, such as those of Bessel, Butterworth and Chebychev.

SUMMARY

Circuits for time-domain operations

- With the basic inverting configurations of Figure 13.1 simple frequency-selective transfer functions can be realized, such as integrators, differentiators, band-pass filters and circuits with PD, PI and PID-characteristics.
- The complex transfer function of the integrator given in Figure 13.2a is $H = -1/j\omega RC$; in the time domain: $v_o = -(1/RC)\int v_i dt$.
- The effects of bias currents and offset voltage in the integrator circuit can be reduced by:

- - having compensation techniques similar to those implemented with voltage amplifiers;
- - restricting the integrating range by adding a resistor in parallel to the capacitor.
- The complex transfer function of the differentiator in Figure 13.5 is given as $H = -j\omega RC$; in the time domain: $v_o = -RC(dv_i/dt)$.
- If the differentiator is to function properly its range must be limited, for instance by adding a resistor in series with the capacitor.

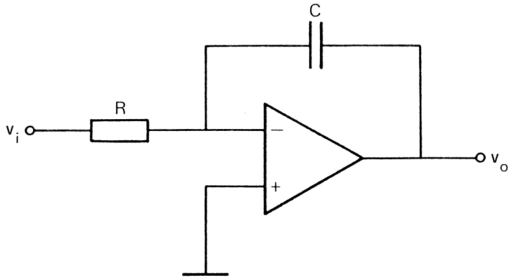
Circuits with high frequency selectivity

- Second-order resonance filters are described using three parameters: the resonance frequency ω_0, the quality factor Q and the transfer at resonance H_0. The quality factor is the ratio between ω_0 and the bandwidth B; $Q = \omega_0/B$.
- Inductorless filters can be created by applying active elements (amplifiers). Two basic configurations are: positive feedback through a passive band-pass network and negative feedback through a notch network. In the first type, the filter parameters are sensitive to variations in component values and the filter may even become unstable. The latter type has guaranteed stability and is less sensitive to component variations.
- Inductorless filters with Bessel, Butterworth and Chebychev characteristics can be realized using operational amplifiers. Such filters are designed with the aid of tables and plots that give the ratio of the component values at the given orders and configurations.

EXERCISES

Circuits for time-domain operations

13.1 In the integrator circuit given below the component values are $C = 1$ μF and $R = 10$ kΩ. The specifications of the operational amplifier are: $|V_{off}| < 0.1$ mV and $|I_{bias}| < 10$ nA. The input is supposed to be zero. At $t = 0$ the output voltage $v_o = 0$. What is the value of v_o after 10 seconds?

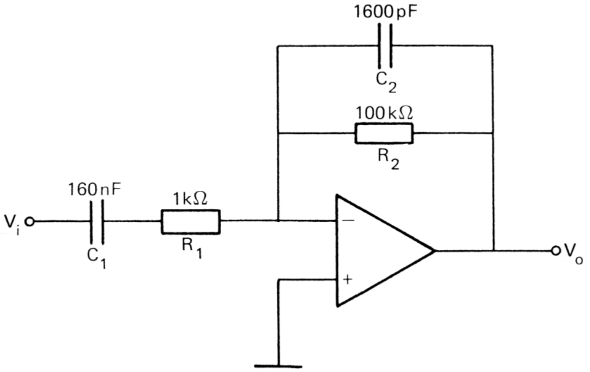


13.2 In the circuit above, resistor R is connected between the ground and the non-inverting input of the operational amplifier. What is v_o at $t = 10$ seconds, under the same conditions as those presented in the preceding question?

13.3 The same circuit is now extended using a resistor $R_0 = 1$ MΩ in parallel with C. What is v_o, under the same conditions as before?

13.4 a. Give the Bode plot (amplitude characteristic only) of the circuit below (asymptotic approximation).

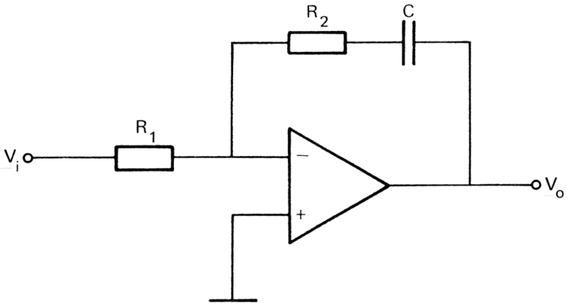
b. The circuit is used as a differentiator, so the phase shift should be 90°. Find the frequency for which the phase shift error is more than 10 degrees (that is: the frequency for which the phase shift is 80°).



13.5 In the next PI-circuit, the component values of R_1, R_2 and C have to be chosen in such a way that the following requirements are satisfied:

- input resistance $R_i > 10$ kΩ
- proportional gain 2
- integration range that goes at least up to 100 Hz.

The operational amplifier may be considered ideal.



Circuits with high frequency selectivity

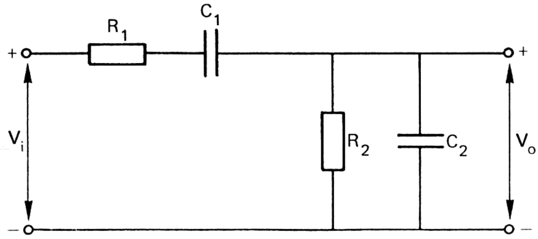
13.6 The transfer function of a certain second-order band-pass filter is given as

$$H(j\omega) = \frac{V_o}{V_i} = \frac{(1 + j\omega\tau)^2}{3 + j\omega\tau/4 - 3\omega^2\tau^2}$$

with $\tau = 10^{-3}$ s.

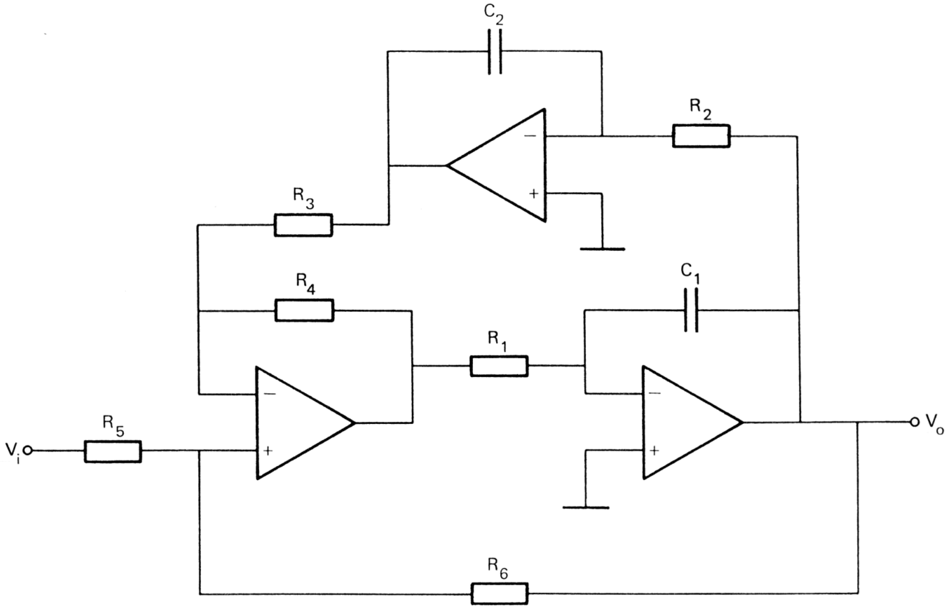
Find the resonance frequency ω_0, the transfer at very low and very high frequencies and the quality factor Q.

13.7 Find the transfer function of the circuit below, and estimate the resonance frequency and the quality factor Q. $\tau = RC = 10$ ms, $R_1 = R_2 = R$ and $C_1 = C_2 = C$.



13.8 Convert the circuit in Figure 13.14 into a second-order high-pass filter with Butterworth characteristics.

13.9 The following circuit is a filter consisting of two integrators and a differential amplifier. This filter is called a dual-integrator loop or a state-variable filter. Find the transfer function. Which components should be made adjustable to vary Q independently of ω_0, and which components should be used to tune ω_0 independently of Q?



14 Nonlinear signal processing with operational amplifiers

This chapter explains how to design circuits for nonlinear analog transfer functions. First we shall discuss the circuits that are used for specific nonlinear operations, such as: the comparator, the Schmitt-trigger and special nonlinear circuits where pn-diodes function as electronic switches. Some specific functions, like logarithmic and exponential functions, and multiplication and division, employ the exponential relationship between the current and the voltage of a pn-diode or bipolar transistor (see Chapters 9 and 10). These electronic systems, which are available as complete integrated circuits, will be discussed in the second part of this chapter, as will various circuits used for arbitrary nonlinear functions.

14.1 Nonlinear transfer functions

This section deals with the circuits that are implemented to realize certain common nonlinear functions. One after another we shall discuss the comparator, the Schmitt-trigger, active voltage limiters and rectifiers.

14.1.1 Voltage comparators

A voltage comparator (or short comparator) is an electronic circuit that responds to a change in the polarity of an applied voltage. The circuit has two inputs and one output (Figure 14.1a). The output has just two levels: high or low, all depending on the polarity of the voltage between the input terminals (Figure 14.1b). The comparator is frequently used to determine the polarity in relation to a reference voltage (Figure 14.1c).

It is possible to use an operational amplifier without feedback as a comparator. The high gain makes the output either maximally positive or maximally negative, depending on the input signal. However, an operational amplifier is rather slow, in particular when it has to return from the saturation state.

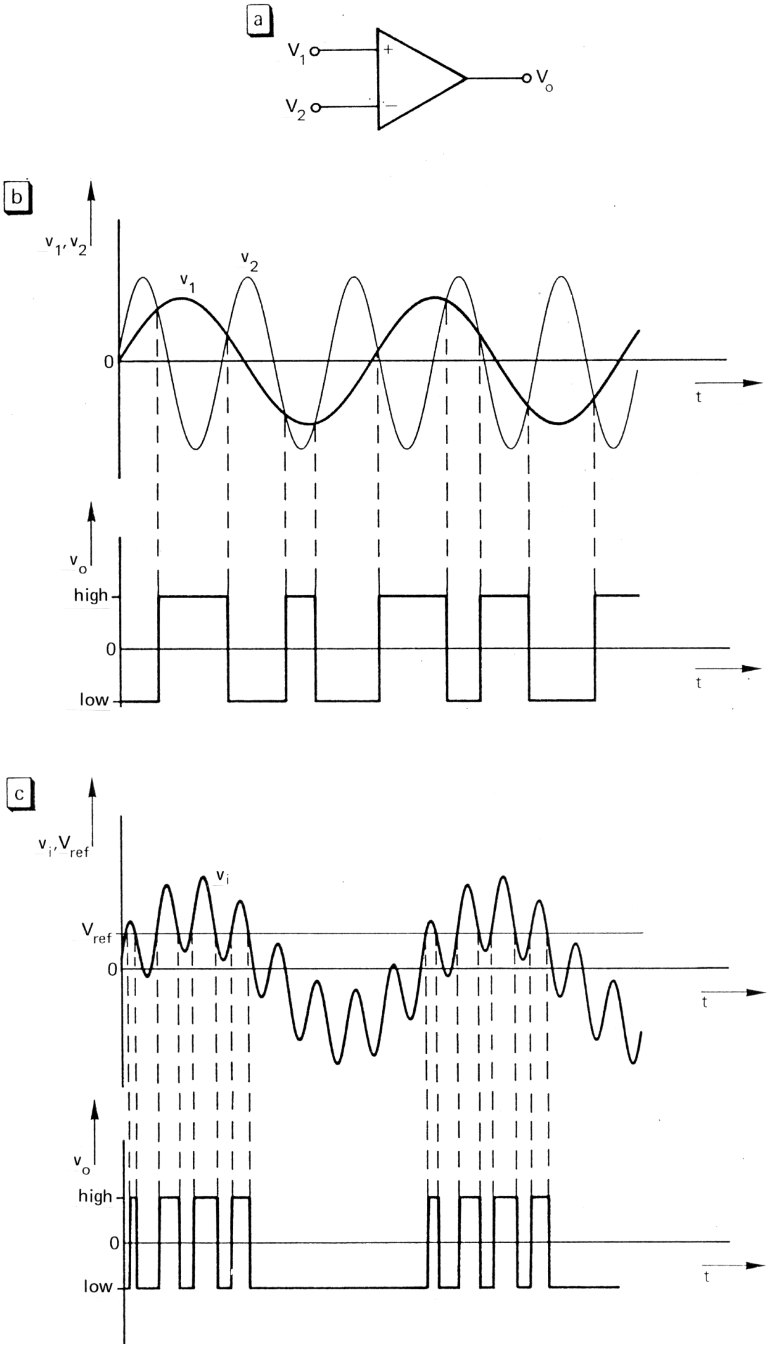

a
V1
V2
+
−
Vo
b
v1, v2
v2
v1
0
t
vo
high
0
low
t
c
vi, Vref
vi
Vref
0
t
vo
high
0
low
t

Figure 14.1. (a) The circuit symbol for a comparator, (b) comparator output voltage v_o as a function of two sinusoidal input voltages v_1 and v_2, (c) the output for an input voltage v_i which is compared to a reference voltage V_{ref}.

Table 14.1. Specifications for two types of comparators: fast (type I) and accurate (type II).

		Type I	Type II
Voltage gain	A	—	$2{\cdot}10^5$
Voltage input offset	V_{off}	±2 mV ± 8μV/K	±1 mV, max. ±4 mV
Input bias current	I_{bias}	5 μA	25 nA, max. 300 nA
Input offset current	I_{off}	0.5 μA ± 7 nA/K	3 nA, max. 100 nA
Input resistance	R_I	17 Ω	–
Output resistance	R_o	100 Ω	80 Ω
Response time	t_r	2 ns	1.3 μs
Output voltage, high	$v_{o,h}$	3 V	–
Output voltage, low	$v_{o,l}$	0,25 V	–

The purpose-designed comparators have a much faster recovery time with response times as low as 10 ns. They have an output level that is compatible with the levels used in digital electronics (0 V and +5 V). Their other properties correspond to a normal operational amplifier and the circuit symbol resembles that of the operational amplifier. Table 14.1 gives the specifications for two different types of comparators, a fast type and an accurate type.

Integrated circuits with two to four comparators on one chip are currently available. Since the power supply pins are combined, the total number of pin connections required for a multi-comparator chip can be kept reasonably low. There are also comparators that have two complementary outputs so that when one output is high, the other is low and vice versa. Other types have an additional control input terminal (strobe) that can switch off the entire circuit. At high strobe input (or low, depending on the type), the comparator output will be high as well, irrespective of the input polarity. At low (or high) strobe input, the circuit operates as a normal comparator.

The comparator not only gives information about the input voltage polarity but also about the moment of polarity change. It is this property that makes the comparator a useful device in various kinds of counter circuits. Imperfections in the comparator (offset, time delay) will cause output uncertainty and much the same goes for input signal noise (Figure 14.2).

The time error increases as the slope of the input signal decreases and noise level increases.

14.1.2 Schmitt-trigger

Noise in the input signal causes fast, irregular comparator output changes (Figure 14.3a). Hysteresis is intentionally introduced to the comparator function (Figure 14.3b) in order to reduce or eliminate output jitter. The output switches from low to high as soon as v_i exceeds the upper reference level V_{ref1}, and from high to low as soon as v_i drops below the lower level V_{ref2}. For proper operation, the hysteresis interval $V_{ref1} - V_{ref2}$ must exceed the noise amplitude. However, great hysteresis will lead to huge timing errors in the output signal.

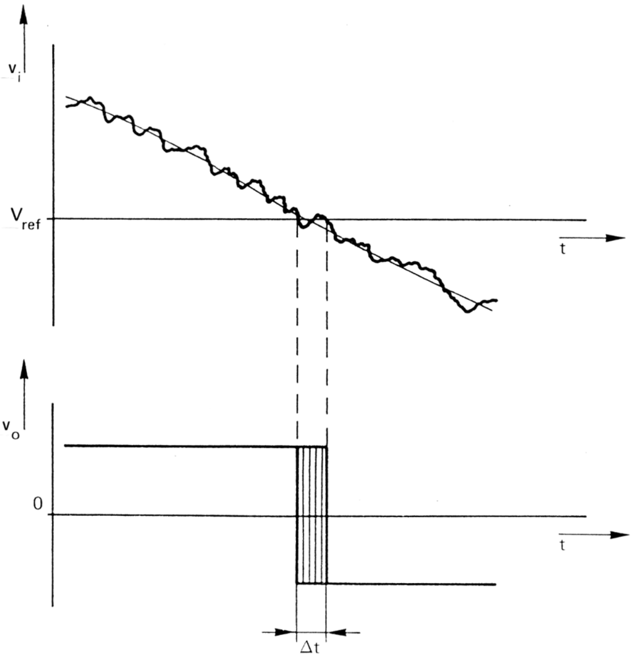


Figure 14.2. Comparator timing uncertainty caused by noise in the input signal.

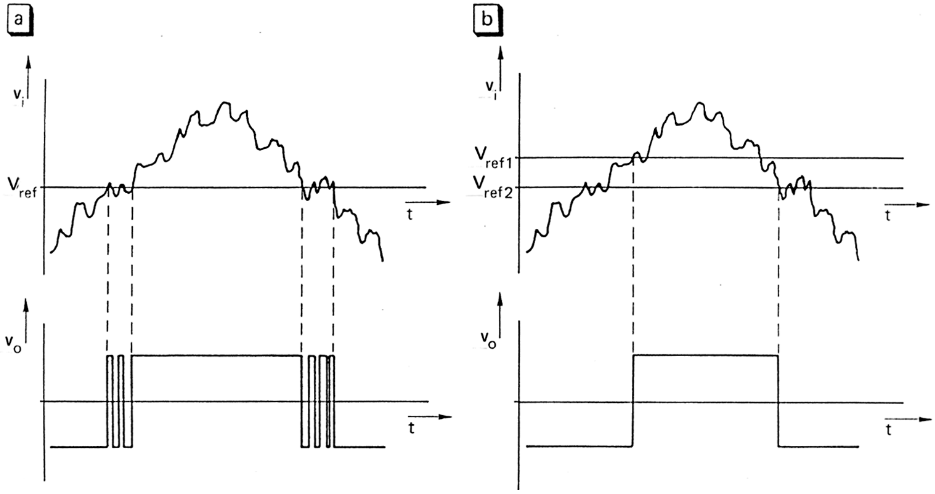


Figure 14.3. (a) Comparator output jitter caused by input noise
(b) reducing jitter by introducing hysteresis.

The right amount of hysteresis is achieved by having an operational amplifier with positive feedback (Figure 14.4). The circuit, which is known as a Schmitt-trigger, operates in the way described below. A fraction β of the output voltage is fed back into the non-inverting input: $\beta = R_1/(R_1 + R_2)$. Suppose that the most positive output voltage is E^+ (usually around the positive power supply voltage) and the most negative output is E^-. The voltage at the non-inverting input will be either βE^+ or βE^-. When v_i is below the voltage on the non-inverting input, then v_o equals E^+ (because of the high gain). This remains a stable situation as long as $v_i < \beta E^+$. If v_i reaches the value βE^+, the output will decrease sharply and so will the non-inverting input voltage. The voltage difference between both input terminals decreases much faster than v_i increases so, within a very short period of time, the output becomes maximally negative (E^-). As long as $v_i > \beta E^-$, the output remains $v_o = E^-$, a new stable state.

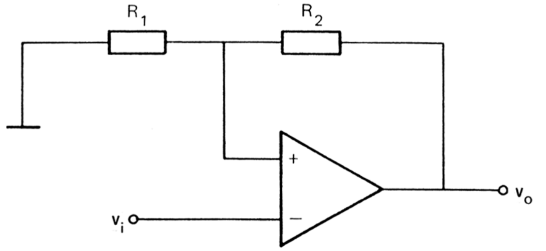


Figure 14.4. A Schmitt-trigger, realized with an operational amplifier and positive feedback (comparator with hysteresis).

The comparator levels of the Schmitt-trigger are apparently βE^+ and βE^-. In conjunction with the positive feedback, even a rather slow operational amplifier can have a fast response time. Evidently, a special comparator circuit is even better.

The switching levels can be varied by connecting R_1 to a reference voltage source V_{ref}. Both levels shift by factor $V_{ref}R_1/(R_1 + R_2)$. In most cases the hysteresis is small compared to the input or output signal, so R_1 will be small compared to R_2. $V_{ref} + (R_1/R_2)v_{o,max}$ and $V_{ref} + (R_1/R_2)v_{o,min}$ approximate the respective switching levels.

14.1.3 Voltage limiters

In Section 9.2.1 we discussed voltage limiters that consist of resistors and diodes. The limit level of such circuits is determined by the diode threshold voltage, which is not well fixed and temperature dependent (about −2.5 mV/K). Those kinds of disadvantages can be overcome by using operational amplifiers. Figure 14.5 shows the basic arrangement of an active limiter with adjustable limit levels and linear range transfer.

This is how it operates: the diodes D_1 and D_2 are either reverse biased (infinite resistance) or conducting (zero resistance with forward voltage V_D). The circuit itself has two stable states. When the input voltage is positive, current will flow through R_1 and D_1 to the amplifier output. Since the amplifier is properly fed back via the conducting diode D_1 the inverting input of the operational amplifier is at ground potential and the output voltage equals $-V_D$, that is: the threshold voltage of D_1. This voltage is sufficiently negative to keep D_2 reverse biased. As there is no current flowing through R_2 the circuit output voltage is zero.

A negative input voltage causes a current to flow from the amplifier output via D_2, R_2 and R_1 to the input. The operational amplifier is then fed back via the conducting diode D_2 but D1 is reverse biased (because v_o is positive). In this state the output voltage equals $v_o = -(R_2/R_1)v_i$. Figure 14.5b shows the resulting transfer characteristic.

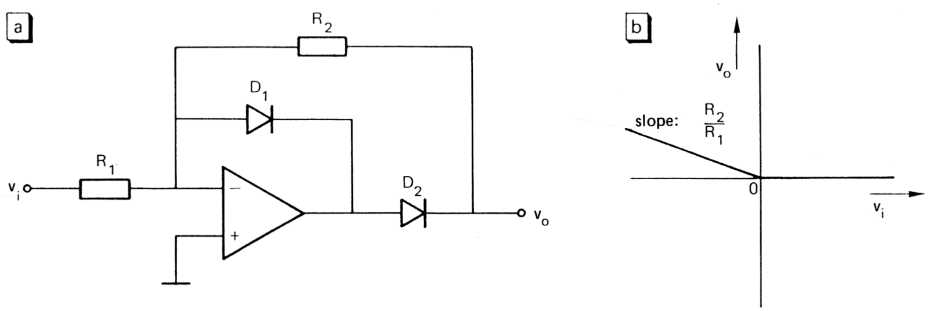


Figure 14.5. (a) An active voltage limiter, (b) the corresponding transfer characteristic.

The circuit in Figure 14.5a limits the output voltage to a minimum value which, in this case, is zero. Conversely, restricting to a maximum value (zero) is achieved by reversing both diodes, thus resulting in the transfer function depicted in Figure 14.6b, curve 1.

Figure 14.6a shows a configuration that allows the characteristic to shift in horizontal and vertical directions. The voltage $V_{ref,1}$ moves the curve in a horizontal direction. As already explained, the current either flows through D_1 or through D_2 and R_2. In accordance with Kirchhoff's rule, this current must be equal to the sum of the currents through R_1 and R_3, hence $V_{ref,1}/R_3 + v_i/R_1$. The breaking point in the characteristic occurs at zero current which means that $v_i = -V_{ref,1}R_1/R_3$ (see Figure 14.6b, curve 2). A vertical shift is obtained using a voltage $V_{ref,2}$ that is connected to the non-inverting input of the operational amplifier (Figure 14.6b, curve 3).

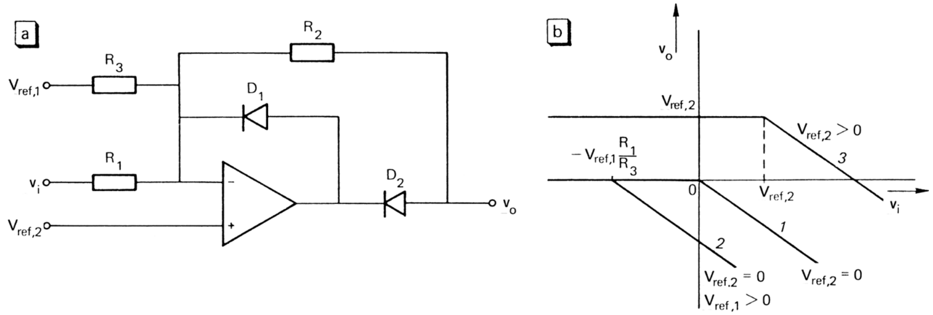


Figure 14.6. (a) A voltage limiter with adjustable limiting levels for the input and output voltage, (b) the corresponding transfer characteristics. Curve 3 applies to non-connected input voltage $V_{ref,1}$ (the floating terminal).

It would appear that the transfer characteristic does not depend on the diode threshold voltages that derive from the high gain of the amplifier. A slight change in the input voltage is enough to switch the diodes from forward to reverse and vice versa. The inaccuracy of the transfer characteristic is only determined by the offset generated by the operational amplifier.

14.1.4 Rectifiers

The aim of a double-sided rectifier is to generate a transfer function that satisfies $y = |x|$. With a single-sided rectifier $y = x$ ($x > 0$) and $y = 0$ ($x < 0$): x and y are voltages or currents.

With passive components the most common double-sided rectifier is the Graetz diode bridge discussed in Section 9.2. It only responds to signals that exceed the threshold voltage of the diodes. Figure 14.7 depicts a simple rectifier circuit without that disadvantage. When the input currents are negative, D_1 is forward biased and D_2 is reverse biased. The non-inverting input voltage of the operational amplifier is zero and so the circuit behaves like the current-to-voltage converter seen in Figure 12.2: $v_o = -i_iR$. When the input currents are positive D_2 is forward biased and D_1 is reverse biased. In that situation the circuit behaves like the buffer amplifier given in Figure 12.4 where the input signal is i_iR and the output signal is the same.

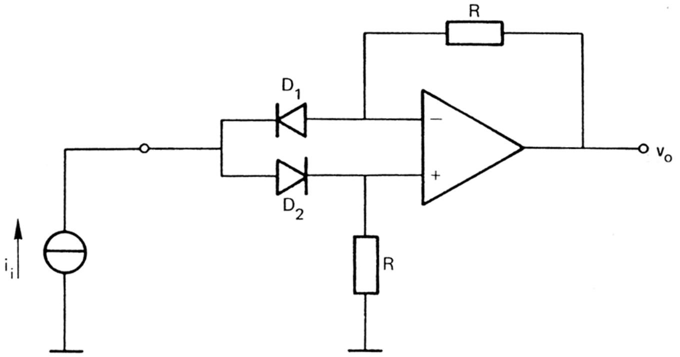


Figure 14.7. A current-voltage converter that operates as a double-sided rectifier.

The circuit depicted in Figure 14.5 can serve as an (inverting) single-sided rectifier. When inversion is not required an additional inverter may be connected to its output. The inverting rectifier can be extended to become a double-sided voltage rectifier by adding twice the input signal to the inverted single-sided rectified signal. This is illustrated in Figure 14.8 where the input voltage is triangular.

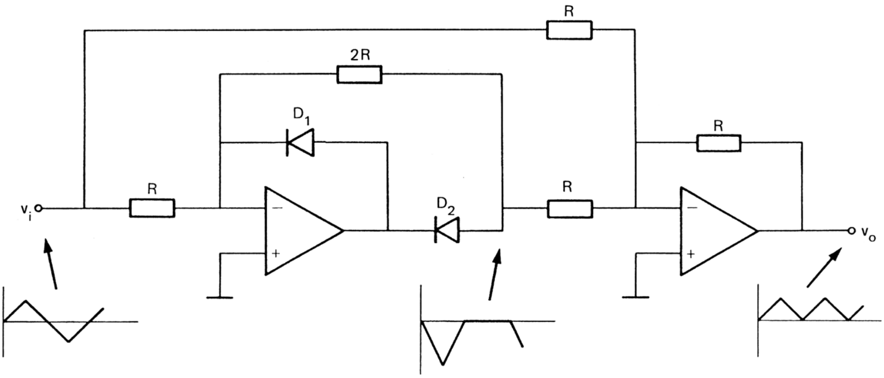


Figure 14.8. A double-sided voltage rectifier, consisting of a single-sided rectifier and a summing circuit.

When working with the circuits described above, the imperfections of the operational amplifier should be borne in mind. It is particularly the limited bandwidth and the slew-rate that could cause problems when signals with relatively high frequencies are being processed.
Alongside of the circuits described in this section, there are a number of other configurations for rectifiers and limiters described in various textbooks on electronics. Some of these circuits have major disadvantages. For instance, since the operational amplifier has no feedback in one of its stable states that substantially slows down its speed. Yet others work only at particular load resistances. So, despite the beautiful simplicity of such circuits they need to be implemented with caution.

14.2 Nonlinear arithmetic operations

14.2.1 Logarithmic converters

Figure 14.9 shows the basic principle of a logarithmic voltage converter. Since the non-inverting input of the ideal operational amplifier is connected to ground, the voltage on the inverted input is zero. The current through R is therefore just v_i/R and it flows entirely through the diode. With this diode, the relation between the voltage and the current is $I_D = I_0 e^{qV_D/kT}$ or $V_D = (kT/q)\ln(I_D/I_0)$. This results in an output voltage equal to:

$$v_o = -V_D = -\frac{kT}{q}\ln\frac{I_D}{I_0} = -\frac{kT}{q}\ln\frac{v_i}{RI_0} = \frac{kT}{q}\cdot\frac{\log(v_i/RI_0)}{\log e} \tag{14.1}$$

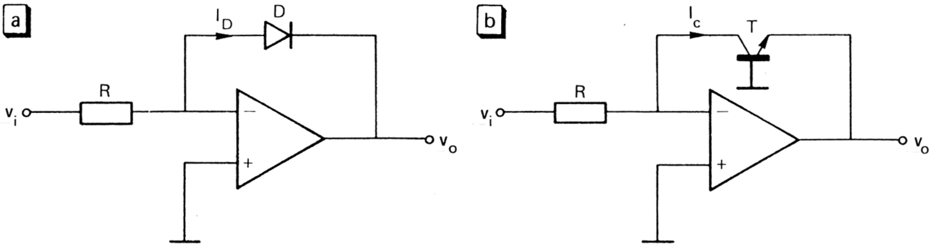


Figure 14.9. The basic circuit for a logarithmic converter, (a) with a pn-diode, (b) with a bipolar transistor.

At room temperature $kT/q \approx 25$ mV and $1/\log e \approx 2.3$ so

$$v_o = -0.06\log\frac{v_i}{RI_0} \tag{14.2}$$

The output voltage decreases by about 60 mV each time the input voltage is increased tenfold. The validity of this exponential relation covers a current range of about 10 nA to several mA. When currents are greater, the series resistance of the diode disturbs the exponential relation. At very low currents the relation is not valid either.

The relation between the collector current I_C and the base-emitter current V_{BE} of a bipolar transistor is also exponential and valid for a much wider range, going from several pA to several mA. Figure 14.9b shows how this bipolar transistor property is exploited in a logarithmic converter. The collector voltage is virtually grounded so the base-collector junction is not forward biased which is one condition for transistor operation in the normal range. Just like the circuit with the diode, the output for this circuit is:

$$v_o = -\frac{kT}{q}\ln\frac{v_i}{RI_0} = -0.06\log\frac{v_i}{RI_0} \tag{14.3}$$

Both circuits can only be used with positive input voltages. To allow in negative input voltages, the polarity of the diode in Figure 14.9a must be reversed and in Figure 14.9b the npn-transistor must be replaced by a pnp-type transistor, however, the circuits will remain unipolar.

The main drawback of the logarithmic converters seen in Figure 14.9 is their tremendous temperature sensitivity. There are two terms that are responsible for this: the term kT/q and the leakage current I_0. The latter can easily be compensated for by a second converter with an identical structure which operates at the same temperature as, for instance, that revealed in Figure 14.10.

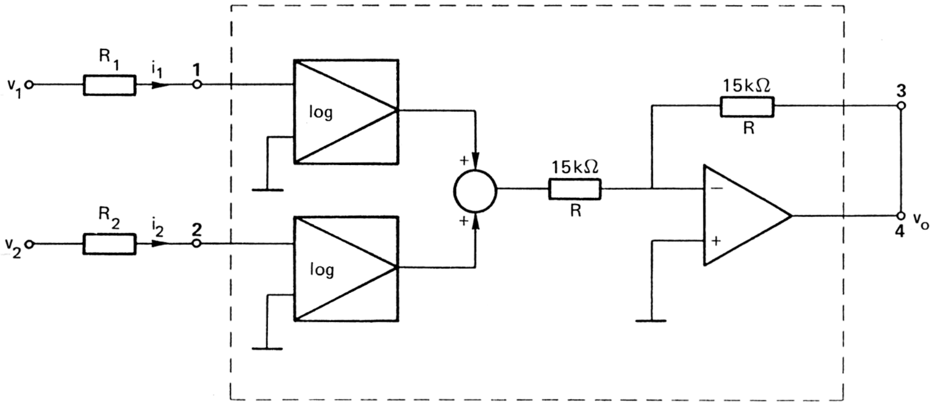


Figure 14.10. The basic layout of an integrated logarithmic converter internally adjusted to base 10.

Subtracting one output from the other will result in a voltage that is proportional to the ratio of the two input signals and that ratio will be independent of I_0 provided that both transistors have the same leakage current. Because of the factor kT/q the remaining temperature coefficient is only 1/300 or 0.33% per K (at room temperature, 300 K).

Other shortcomings are caused by imperfections in the operational amplifier, such as with the bias currents and the offset voltage, both of which contribute to an additional current through the diode or transistor. If one takes into account the bias current and the offset, the total current will amount to:

$$I_D = \frac{v_i}{R} + \left|\frac{V_{off}}{R}\right| + |I_{bias}| \tag{14.4}$$

and the modulus of I_{bias} and V_{off} is taken because we do not know the polarity of these quantities. The output of the diode circuit becomes $v_o = V_{off} - V_D$ (V_D includes the additional current derived from V_{off} and I_{bias}). The output of the transistor circuit amounts to $v_o = -V_{BE}$. Although the collector voltage is $|V_{off}|$, it does not affect the base-emitter voltage. The influence of V_{off} and I_{bias} can be compensated in a similar way to that done in linear amplifier circuits (Section 12.2.2).

Figure 14.10 gives the internal structure of a commercial type logarithmic converter. As indicated in Figure 14.10b, the squares with "log" printed inside them represent circuits but they do not have resistor R. That has to be connected externally by the user. These converters come in two types: "P" and "N", for positive and negative input voltages. Table 14.2 lists the main specifications for such a converter. It is composed of two logarithmic converters, a subtractor (differential amplifier) and an output amplifier. The output voltage is proportional to the logarithm of the ratio between the two input currents I_1 and I_2, which is why it is called a log-ratio converter. Between specific boundaries the transfer (or scaling factor) can be determined by the system user simply by placing an external resistor between pins 3 and 4. The low input impedance of the integrated circuit also allows input voltage conversion to take place if external resistors R_1 and R_2 are connected in series with the inputs.

Table 14.2. The specifications of a log-ratio converter.

Transfer function	$V_o = -K \log I_1/I_2$
Input current range	I_1, I_2; (+/–) 1 nA … 1 mA
Scaling factor	1 V decade ±1% ± 0.04%/K
Bias current:	$I_{bias,1}$, $I_{bias,2}$: 10 pA (doubles per 10 K)
Input offset voltage	$V_{off,i} = \pm$ 1 mV (max.) ± 25 μV/K
Output offset voltage	$V_{off,o} = \pm$ 15 mV (max., adjustable to zero) ± 0.3 mV/K

Example 14.1

Voltage v_i should be converted into voltage $v_o = -2\log(v_i/V_{ref})$ using the converter shown in Figure 14.10 and the corresponding specifications in Table 14.2. Voltage v_i ranges from 10 mV to 10 V and a reference voltage of 10 V is available.

To minimize errors emanating from bias currents and offset voltages the currents are made as large as possible so that for instance $I_1 = I_2 = 1$ mA. This is achieved with $R_1 = R_2 = 10\ k\Omega$. The scaling factor is set at 2 (i.e. twice the nominal value) by connecting a resistor of 15 kΩ between terminals 3 and 4. At the lowest input voltage, the output is $-2\log 10^{-3} = 6$ V. The output error at that input voltage point consists of two parts (Table 14.2), one arising from $V_{off,o}$: $2 \times (\pm 15)$ mV (the gain is twice the nominal value) and the other arising from $V_{off,i}$: $v_o = -2\log\{(10\ mV \pm 1\ mV)/10\ V\}$, a value which lies between +5.98 and +6.09 V. There is therefore a maximum error of about 90 mV. Compared to these offset errors other errors are negligible.

14.2.2 Exponential converters

The pn-diode and the bipolar transistor can also be used to create an exponential converter (sometimes also known as an anti-log converter) as shown in Figure 14.11.

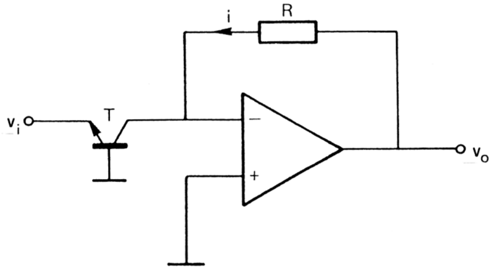


Figure 14.11. The basic circuit of an exponential voltage converter.

Assuming that the operational amplifier is ideal, the output voltage will equal $v_o = IR = I_0 R e^{-qv_i/kT}$. The minus sign derives from the fact that $v_o = V_{EB} = -V_{BE}$. The circuit only operates with negative input voltages. A circuit that has a pnp-transistor rather than an npn-transistor will only operate with positive input voltages.

The great temperature sensitivity of I_0 can be compensated by using the same method as that used in the logarithmic converter, in other words, by adding a second, identical circuit. The remaining temperature coefficient deriving from the factor kT/q is sometimes reduced by a built-in temperature compensation circuit. The transfer of such an exponential converter can be given as $v_o = -V_{ref} e^{-v_i/K}$, where V_{ref} is an internal or external reference voltage and K a scaling factor.

Depending on the external connections, some of the types available can also be used as logarithmic converters. Figure 14.12 demonstrates that possibility. The system consists of an exponential converter (having the transfer mentioned above), a reference voltage V_{ref} and an operational amplifier. In Figure 14.12a, the device is connected as an exponential converter and in Figure 14.12b as a logarithmic converter. The transfer in the exp-mode is $v_o = -v_y = V_{ref} e^{v_i/K}$ and in the log-mode it is $v_o = v_x$ and $v_i = -v_y$, so $v_o = -K \ln(v_i/V_{ref})$. The operational amplifiers are presumed to be ideal.

14.2.3 Multipliers

Most commercial analog multipliers are based on a combination of logarithmic and exponential transfer functions. Figure 14.13 provides the functional diagram of a multiplier composed of the exponential and logarithmic converters seen in Figures 14.9 and 14.11. The transfer for logarithmic converters is $v_{o,j} = -(kT/q)\ln(v_{i,j}/I_0 R_j)$, $j = 1,2,3$, whereas for the exponential converter it is $v_{o,4} = I_0 R_4 e^{-qv_{i4}/KT}$. The voltage after summation equals $-(kT/q)\ln(v_{i,1} v_{i,2} R_3 / v_{i,3} I_0 R_1 R_2)$, so the output voltage is

$$v_{o,4} = \frac{v_{i,1} v_{i,2}}{v_{i,3}} \cdot \frac{R_3 R_4}{R_1 R_2} \tag{14.5}$$

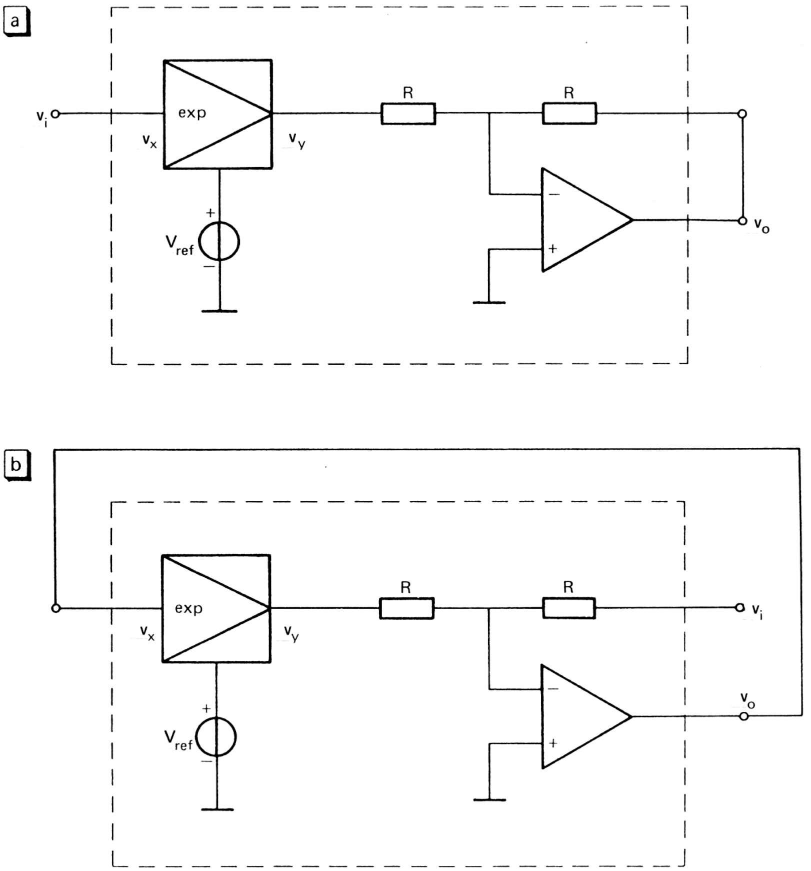

a
b
exp
v_i
v_x
v_y
R
R
V_{ref}
+
−
v_o

Figure 14.12. An integrated exponential converter, (a) connected as an exponential converter, (b) connected as a logarithmic converter.

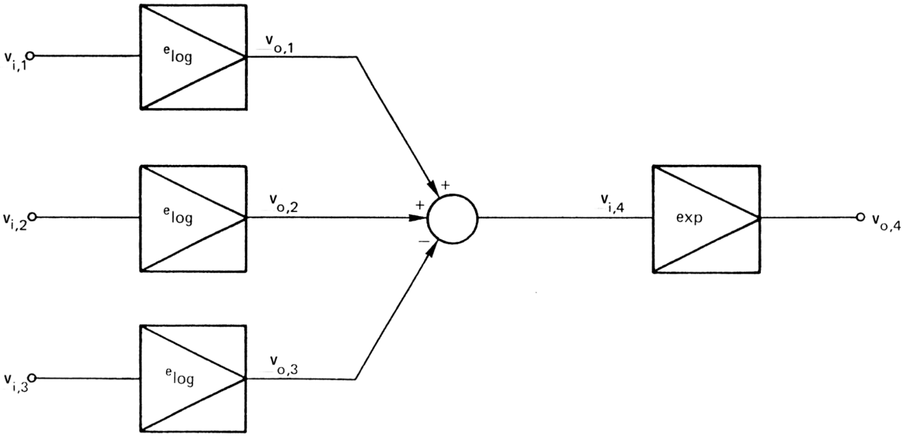

$v_{i,1}$
$v_{i,2}$
$v_{i,3}$
^{e}log
^{e}log
^{e}log
$v_{o,1}$
$v_{o,2}$
$v_{o,3}$
+
+
−
$v_{i,4}$
exp
$v_{o,4}$

Figure 14.13. A functional diagram of an analog multiplier based on exponential and logarithmic converters and a summing circuit.

Despite the heavy temperature dependence of the individual converters, the output does not depend on the leakage current I_0 and the temperature T (which provides identical converters and the same temperature). Another advantage is that the circuit can act as a divider. The main disadvantage is the unipolarity: multiplication is only performed in one quadrant.

To dispel this shortcoming, multipliers are constructed that are capable of handling both polarities. Such multipliers are also based on the logarithmic and exponential relations of a bipolar transistor and are available as integrated circuits.

Table 14.3 is part of the specification list for a low-cost analog multiplier integrated circuit. This integrated device has three inputs: X, Y and Z. The function of the Z input will be discussed in Section 14.2.4. The offset voltages of the three inputs can be individually adjusted to zero by connecting compensation voltages to three extra terminals. In the data sheet the manufacturer specifies how this can be done. The crosstalk in the specification list is the output voltage derived from just one input voltage while all the other input voltages are zero. The non-linearity of the device is specified for the individual X and Y channels at maximum input voltage in both channels.

Table 14.3. The specifications for an analog multiplier.

Transfer function	$V_o = Kv_xv_y$; $K = 0.1\ \text{V}^{-1}$
Scale error	±2% ± 0.04%/K
Max. input voltages	± 10 V
Non-linearity	$v_x = v_o = 20$ V (peak-peak); ± 0.8%
	$v_y = v_o = 20$ V (peak-peak); ± 0.3%
Crosstalk (peak-peak value 50Hz)	$v_x = 20$ V, $v_y = 0$; $v_o < 150$ mV
	$v_y = 20$ V, $v_x = 0$; $v_o < 200$ mV
Input resistance X-input	10 MΩ
Y-input	6 MΩ
Z-input	36 kΩ
Output resistance	100 Ω
Output offset voltage	adjustable to zero
t.c. of the offset voltage	±0.7 mV/K
Bias current X,Y-input	3 μA
Z-input	±25 μA
Bandwidth	750 kHz

14.2.4 Other arithmetic operations

The multiplier discussed in the preceding section can also be applied to division and used to realize square power and square root transfer functions. A quadratic transfer function is achieved by simply connecting the input signal to both multiplier inputs: $v_o = Kv_i^2$. The division and extraction of the square root is obtained by introducing a feedback configuration and using an operational amplifier. Most commercial devices have a built-in amplifier for that particular purpose.

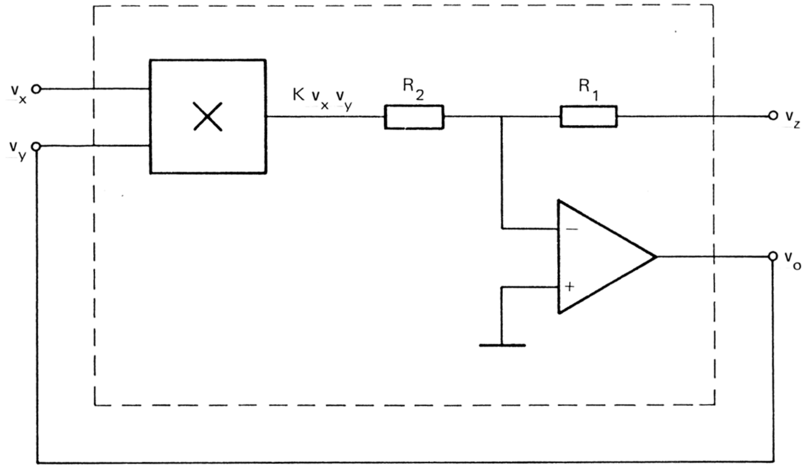


Figure 14.14. An example of a divider circuit: $v_o = -(R_2/KR_1)\cdot(v_z/v_x)$. The multiplier is in the feedback loop of the operational amplifier. The IC can also be connected as a multiplier and as a square rooter.

Figure 14.14 shows the functional structure of a divider circuit, it also shows the *Z*-input from the foregoing section. With proper feedback, the inverting input voltage of the operational amplifier is zero so $v_zR_2/(R_1 + R_2) + Kv_xv_yR_1/(R_1 + R_2) = 0$. Furthermore $v_y = v_o$, so $v_o = -v_zR_2/Kv_xR_1$ and the output is proportional to the ratio of v_z and v_x. The proportionality factor can freely be chosen with R_1 and R_2. The divider only operates for positive v_x values. In the case of negative values the negative feedback changes into positive feedback thus destabilizing the system. The division operation error margin depends on the errors made by the multiplier and the operational amplifier. If one assumes that the multiplier output offset voltage is $V_{off,a}$ and that the input offset voltage of the operational amplifier is $V_{off,b}$ then:

$$\frac{\left(Kv_xv_y + V_{off,a}\right)R_1}{R_1 + R_2} + \frac{v_zR_2}{R_1 + R_2} = V_{off,b} \tag{14.6}$$

and thus, with $v_o = v_y$:

$$v_o = -\frac{v_zR_2}{Kv_xR_1} + \frac{V_{off,b}\left(R_1 + R_2\right)}{Kv_xR_1} - \frac{V_{off,a}}{Kv_x} \tag{14.7}$$

This shows that the relative error in v_o can be very large at low v_x values.
The circuit given in Figure 14.14 can be used to realize a square root transfer function. The *X* and *Y* inputs are connected to each other. Since $v_x = v_y = v_o$, the output voltage is $v_o = -R_2v_z/Kv_oR_1$, or: $v_o = \surd(-R_2v_z/KR_1)$ Evidently, v_z can only be negative.

Transfer functions with powers other than 2 or $\frac{1}{2}$ can be realized with combinations of logarithmic and exponential converters. The general principle is given in Figure

14.15. There m is an amplifier ($m > 1$) or an attenuator ($m < 1$). With $m = 2$ and $m = \frac{1}{2}$, the transfer corresponds respectively to a square power and a square root. These circuits are more accurate than those with analog multipliers. A disadvantage is the higher complexity and the unipolarity.

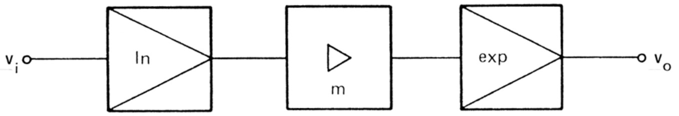


Figure 14.15. An arbitrary power function is realized using a logarithmic converter, an amplifier/attenuator and an exponential converter: $v_o = Kv_i^m$.

Example 14.2

The mass flow Φ_m of a gas satisfies the equation $\Phi_m = F\sqrt{(P\Delta p/T)}$, if P is the total pressure, Δp the pressure difference across the flow meter and T the absolute temperature. F is a system constant. The quantities P, Δp and T can be measured using electronic sensors. Let us assume that the output signals for those sensors are x_P, $x_{\Delta p}$ and x_T, respectively. An electronic circuit used to determine Φ_m is shown in Figure 14.16. The operation follows the steps given in this figure. The output signal is $x_o = \sqrt{(x_P x_{\Delta p}/x_T x_{ref})}$.

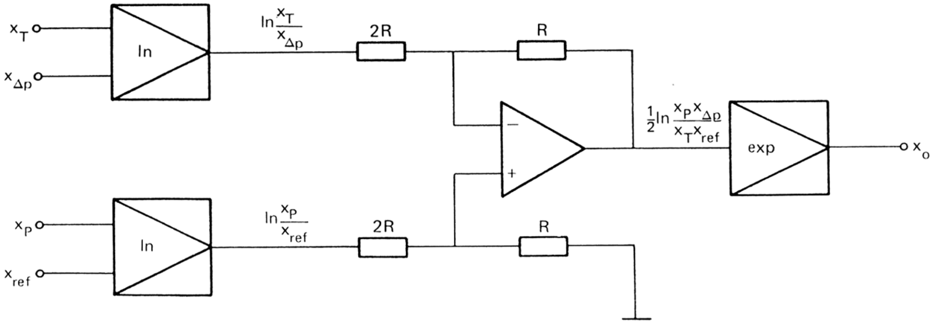

x_T
$x_{\Delta p}$
ln
$\ln\frac{x_T}{x_{\Delta p}}$
2R
R
$\frac{1}{2}\ln\frac{x_P x_{\Delta p}}{x_T x_{ref}}$
exp
x_o
x_P
x_{ref}
ln
$\ln\frac{x_P}{x_{ref}}$
2R
R

Figure 14.16. An analog electronic circuit used to determine $\phi_m = F\sqrt{P\Delta p / T}$; the logarithmic converters are the same type as those given in Figure 14.10.

14.2.5 A piecewise linear approximation of arbitrary transfer functions

An arbitrary transfer function $y = f(x)$ can be approximated using segments of straight lines (Figure 14.17). The smaller the segments, or the more segments there are within the interval $[x_{min}, x_{max}]$, the better the approximation will be.

The approximation is obtained from a combination of similarly shaped elementary functions (Figure 14.17b), in which the transfer function of the limiter from Section 4.2.3 can be recognized. To simplify the explanation we shall take that circuit as the basis for the segmented approximation.

From Figure 14.17b it is clear that it is the weighted addition of the elementary functions that results in the segmented function. For each of these sub-functions we need to take the circuit as depicted Figure 14.14a with the fixed values of R_1, R_2 and R_3, and the adjustable reference voltages. The addition is done by means of an operational amplifier along the lines of the method shown in Figure 12.3b. If necessary a normal linear transfer can be added as well. Figure 14.18 provides an example of such a configuration with a corresponding transfer characteristic that contains two break points (or three segments).

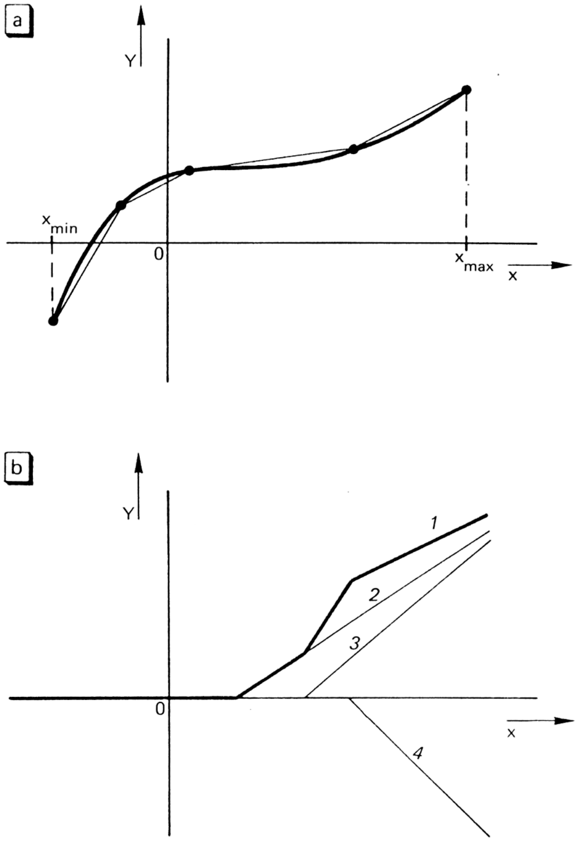


Figure 14.17. (a) Piecewise linear approximation at a number of segments, (b) an example of function 1, composed of the elementary functions 2, 3 and 4.

The reference voltages $V_{ref,j}$ determine the position of the break points. With resistances $R_{i,j}$ the slope of the characteristic can be adjusted from the j-th break point. The slope $-R_t/R_{i,j}$ of the linear function is reduced at each following break point by a quantity $-R_t/R_{i,j}$ times the (negative) slope of the j-th limiter. The whole characteristic can be rotated around the origin by varying R_t.

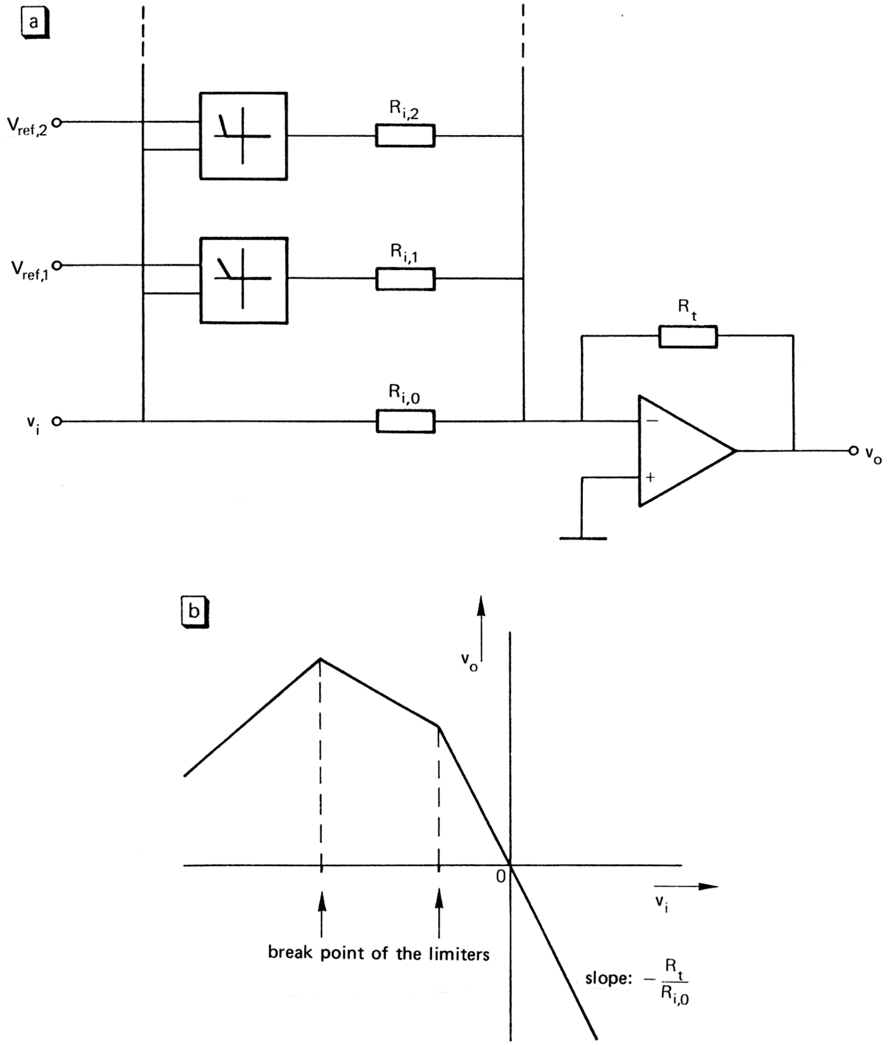

a
$V_{ref,2}$
$R_{i,2}$
$V_{ref,1}$
$R_{i,1}$
R_t
$R_{i,0}$
v_i
−
+
v_o
b
v_o
0
v_i
break point of the limiters
slope: $-\frac{R_t}{R_{i,0}}$

Figure 14.18. (a) A circuit configuration for the weighted summation of the elementary transfer functions seen in Figure 14.6b, using the circuits that feature in Figure 14.5a. (b) the piecewise linear approximation of a non-linear transfer function.

SUMMARY

Nonlinear transfer functions

- A comparator is a differential amplifier with a very high gain and a fast response. Its output is either high or low, depending on the polarity of the input voltage.
- A Schmitt-trigger is a comparator with predetermined hysteresis. Like the comparator, it is used to determine the sign (polarity) of a voltage.

- The Schmitt-trigger has better noise immunity due to its hysteresis but this also introduces an additional time or amplitude error when two signals are compared.
- A Schmitt-trigger can be realized by using an operational amplifier that provides positive feedback.
- In active limiters and rectifiers consisting of diodes and operational amplifiers, the switching levels are independent of the diode forward voltage.

Nonlinear arithmetic operations

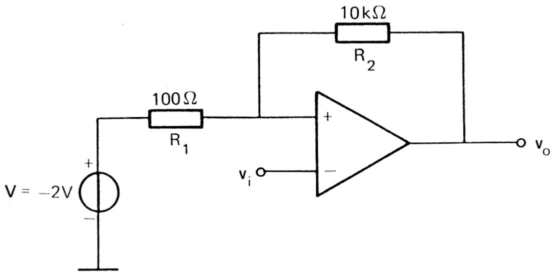
- Logarithmic and exponential signal converters are based on the exponential relationship between the current and the voltage of a pn-diode or on the even more accurate exponential relationship between the collector current and the base-emitter voltage of a bipolar transistor.
- The inherent great temperature sensitivity of a logarithmic and exponential converter (which is due to the leakage current) is eliminated by compensating with a second, identical diode or transistor.
- Most analog signal multipliers are based on a combination of logarithmic and exponential converters. They are available as single-quadrant or four-quadrant multipliers.
- Analog multipliers can also be used for division and square rooting. It is for that very purpose that many integrated multipliers have built-in circuits and can be employed by the user if the appropriate external connections are made.
- When using analog multipliers, special attention should be given to the imperfections of the integrated circuits, such as voltage offset and nonlinearity.
- Arbitrary nonlinear transfer functions can be realized through piecewise linear approximation, built up from a combination of limiters and a summing circuit.

EXERCISES

Nonlinear transfer functions

14.1 A voltage v_g from a source with source resistance $R_g = 1\ \text{k}\Omega$ is compared with a reference voltage V_{ref} that has an inaccuracy of ±2 mV. A comparator with specifications as listed in Table 14.1, type I is used. Calculate the total inaccuracy of this comparison in mV, over a temperature range that goes from 0 to 70 °C.

14.2 Find the switching levels and the hysteresis of the Schmitt-trigger in the figure below. The operational amplifier has limited output voltages of −15 and +15V.

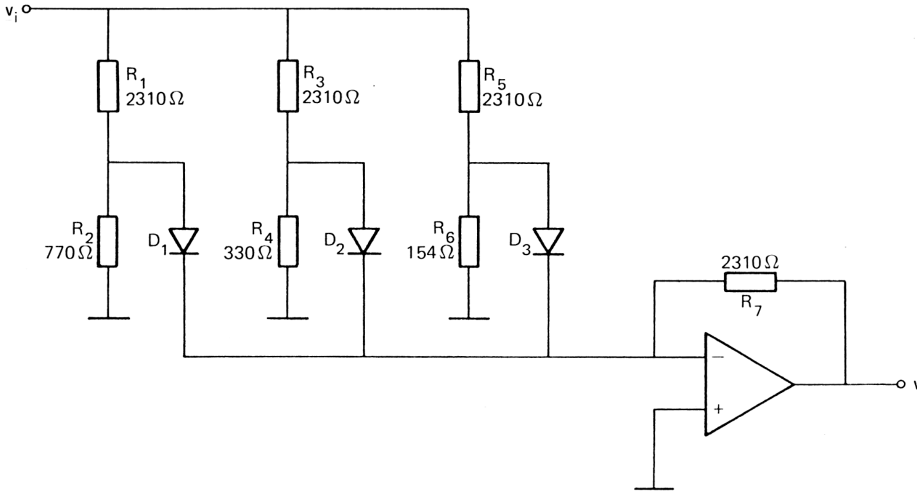


14.3 Draw, in one figure, the output voltage, the non-inverting input voltage and the input voltage of the circuit seen in exercise 14.2. Assume that the input voltage is triangular, the amplitude 5 V and the mean value zero.

14.4 The component values of the circuits seen in Figure 14.6 are $R_1 = R_2 = \frac{1}{2}R_3 = 10\ \mathrm{k\Omega}$; $V_{ref,1} = -3$ V, $V_{ref,2} = 2$ V. Make a sketch of the transfer v_o/v_i.

14.5 Draw the voltages at the points indicated by the arrows in Figure 14.8 showing what it would be like if all the diodes were reverse connected while the input signal remained the same.

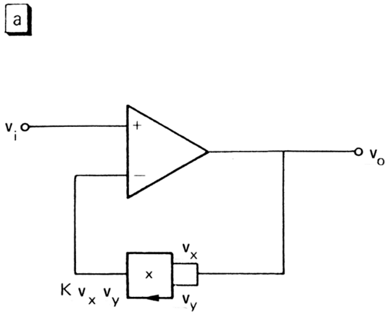
14.6 Find the transfer function of the circuit given below. Make a sketch of the transfer function v_o/v_i. The forward voltage of the diodes is 0.5 V.

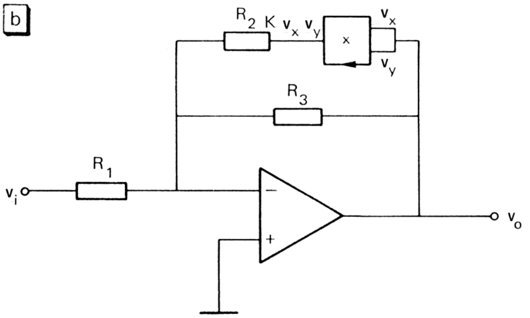


Nonlinear arithmetic operations

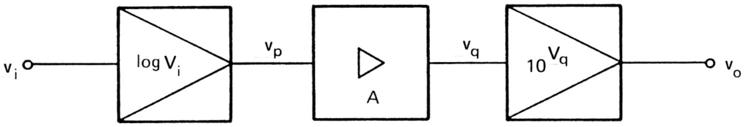
14.7 Logarithmic converters with transfer $v_o = K_L \ln(v_i/V_L)$ and exponential converters with transfer $v_o = K_E e^{v_i/V_E}$ are termed complementary if the total transfer of the converters in series equals 1. Find the conditions for this with respect to K_L, K_E, V_L and V_E.

14.8 Find the transfer functions v_o/v_i for the two circuits given below. The scaling factor is $K = 1$ and the operational amplifier has ideal properties.





14.9 Find the transfer function for the following system. The transfer for the log-converter is $v_p = -\log(v_i/10)$ while the transfer for the exponential converter is $v_o = -10^{-v_q/10}$; $A = 5$.



14.10 Design a circuit with logarithmic and exponential converters that only has one operational amplifier and which realizes the function

$$v_o = \frac{v_3}{v_1^{1/3} \cdot v_2^{2/3}}$$

14.11 The bias currents of a log-ratio converter with transfer $K\log(I_1/I_2)$ are 1 nA at their maximum. $K = 1$ V. Calculate the output voltage for $I_1 = I_2 = 1$ μA. What is the error in the output for $I_1 = 10$ μA and $I_2 = 100$ μA?

15 Electronic switching circuits

The subject of this chapter is electronic switches and circuits composed of electronic switches. In the first part we shall be discussing the general properties of electronic switches and introducing various components that can function as electronically controllable switches. In the latter part of the chapter, time multiplexers and sample-hold circuits – the two types of circuits that make use of electronic switches – will be described. Both kinds of circuits are widely used in instrumentation.

15.1 Electronic switches

There are many components that can serve as switches, some of which – such as diodes and transistors – have already been discussed in preceding chapters. In this section we shall be looking particularly at their properties as electronic switches. The components we shall consider are these: reed switches, photo-resistors, pn-diodes, bipolar transistors, junction field-effect transistors, MOSFETs and thyristors. First, though, something must be said about the general properties of electronic switches.

15.1.1 The properties of electronic switches

When on, the ideal switch forms a perfect short-circuit between two terminals (see 1 and 2 in Figure 15.1a) and when off a perfect means of isolation. The ideal switch has zero response time (it switches on directly as required). Finally, the control terminal (3) is isolated from the circuit terminals which are, in turn, isolated from ground (to form a so-called floating switch). Obviously, an actual switch only partially meets this ideal behavior. The main imperfections are depicted in Figure 15.1b. but even a model as complicated as this does not take into account all the properties of a switch. Usually the dynamic properties of a switch are specified in a time diagram. Figure 15.2 clarifies the commonly used specifications with respect to dynamic behavior. The quantity x_i is connected to one terminal (input), x_o is the quantity at the other terminal (output). The quantity to be switched on and off may be a voltage or a current. In practice, two types of switches are distinguished: voltage switches and current switches. This distinction is based on how the switch is configured in an electronic circuit, rather than on its construction. For instance, it is more appropriate to use a switch with a large offset voltage V_{off} as a current switch than as a voltage switch.

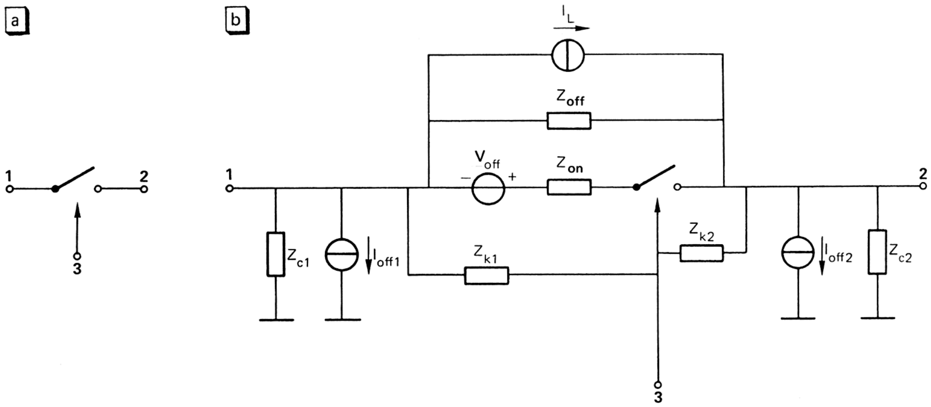


Figure 15.1. (a) A circuit symbol of an electronically controllable switch, (b) a model showing some of the imperfections of an actual switch.

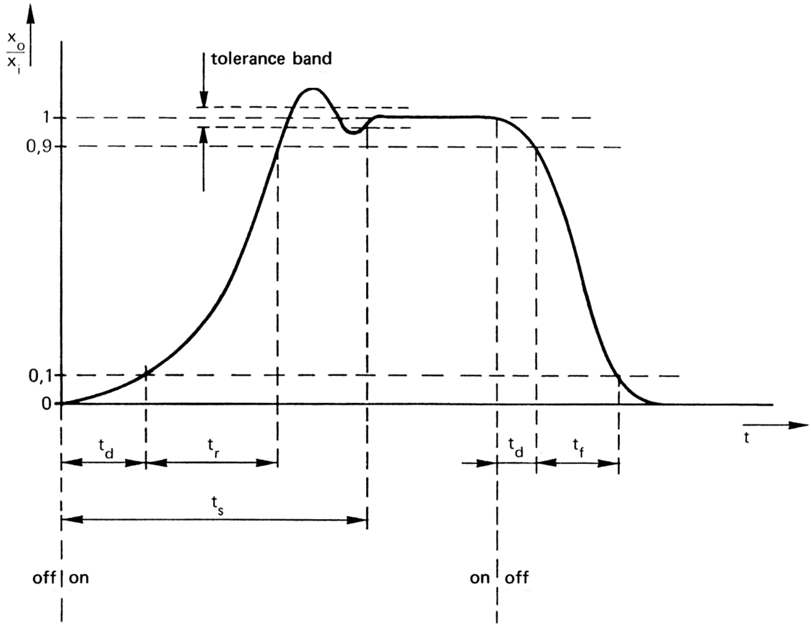


Figure 15.2. The dynamic properties of a switch, in terms of time delays. t_d(on) = turn-on delay time, t_r = rise time, t_s = settling time (the space of time between the on-command and the output within a specified error band around the steady-state), t_d(off) = turn-off delay time, t_f = fall time.

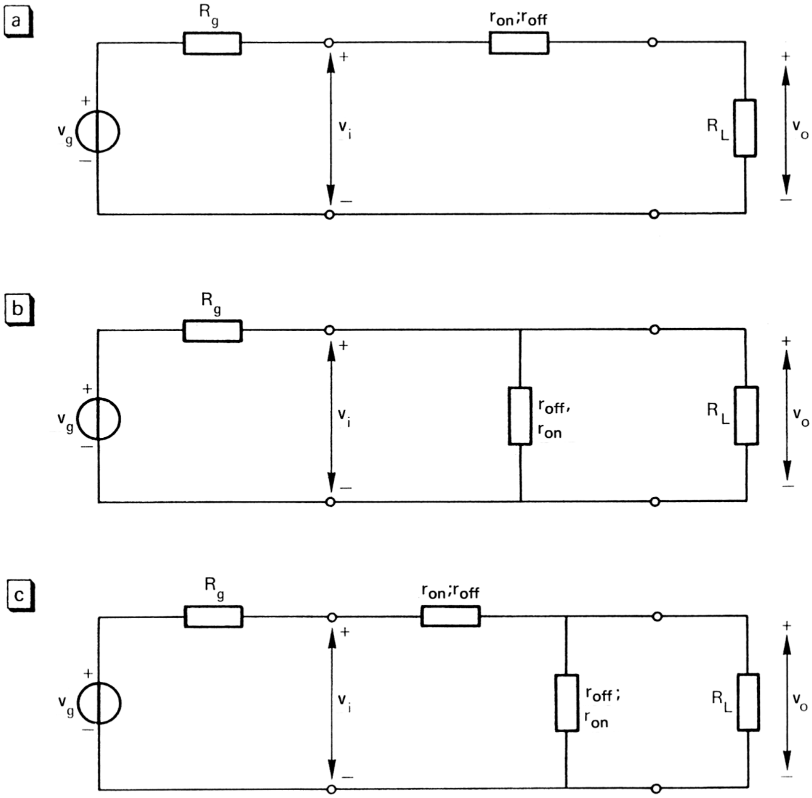


Figure 15.3. Voltage switching with (a) a series switch, (b) a shunt switch, (c) a series-shunt switch.

The salient imperfections of an electronic switch are its on-resistance and its off-resistance. Together with the connected circuits these imperfections may well lead to serious transfer errors. The three different configurations shown in Figure 15.3 reveal how a voltage from a signal source with source resistance R_g is switched to a load R_L (i.e the input resistance of the connected system).

For each of these configurations the transfer in both the on-state and the off-state can be calculated using the formula for the voltage divider.

* **series switch** (Figure 15.3a).

$$\text{On:} \quad \frac{v_o}{v_g} = \frac{R_L}{R_g + r_{on} + R_L} \tag{15.1}$$

$$\text{Off:} \quad \frac{v_o}{v_g} = \frac{R_L}{R_g + r_{off} + R_L} \tag{15.2}$$

In the perfect switch the transfer in the on-state should be equal to $R_L/(R_g + R_L)$ (thus corresponding to a voltage divider configuration). For optimal transfer in the on-state the on-resistance of the switch must satisfy $r_{on} << R_g + R_L$. Usually, $R_g << R_L$, so the condition reduces to $r_{on} << R_L$. The ideal transfer in the off-state is zero which means that the condition for the switch is $r_{off} >> R_L$.

* **shunt switch** (Figure 15.3b).

$$\text{On:}\quad \frac{v_o}{v_g} = \frac{R_L//r_{off}}{R_g + R_L//r_{off}} = \frac{R_L}{R_g R_L/r_{off} + R_g + R_L} \tag{15.3}$$

$$\text{Off:}\quad \frac{v_o}{v_g} = \frac{R_L//r_{on}}{R_g + R_L//r_{on}} = \frac{r_{on}}{r_{on}\left(1 + R_g/R_L\right) + R_g} \tag{15.4}$$

To estimate the ideal transfer the on-resistance must satisfy $R_g R_L/r_{off} << R_g + R_L$ or, as it is usually so that $R_g << R_L$, the requirement will become $r_{off} >> R_g$. For zero transfer in the off-state: $r_{on}(1 + R_g/R_L) << R_g$ or, since $R_g << R_L$: $r_{on} << R_g$.

* **series-shunt switch** (Figure 15.3c).

$$\text{On:}\quad \frac{v_o}{v_g} = \frac{R_L//r_{off}}{R_g + +r_{on} + R_L//r_{off}} = \frac{R_L}{R_g R_L/r_{off} + r_{on} R_L/r_{off} + r_{on} + R_g + R_L} \tag{15.5}$$

$$\text{Off:}\quad \frac{v_o}{v_g} = \frac{R_L//r_{on}}{R_g + r_{off} + R_L//r_{on}} = \frac{r_{on}}{\left(1 + r_{on}/R_L\right)\left(r_{off} + R_g\right) + r_{on}} \tag{15.6}$$

For the same reasons as before, the requirements for the on and off-resistances are: $r_{off} >> R_g$; $r_{off} >> r_{on}$ and $r_{on} << R_L$. Under the conditions $r_{on} << R_L$ and $r_{off} >> R_g$, the off-state transfer is approximately equal to r_{on}/r_{off}.

These requirements are summarized in Table 15.1. In most cases, R_L is large and R_g is small. This interferes with the requirement $r_{off} >> R_L$ for the series switch and $r_{on} << R_g$ for the shunt switch. This problem does not apply to the series-shunt switch, there the only requirement is $r_{off} >> r_{on}$, regardless of the source and load which can easily be met by most switch types. A disadvantage of the series-shunt switch is that there is a need for two complementary switches or two equal switches with complementary control signals. In particular applications, however, this can even be an advantage, as will be explained in Section 15.2.

Table 15.1. The requirements for the on and off resistances for the three configurations of Figure 15.3.

	r_{on}	r_{off}
Series	$\ll R_L$	$\gg R_L$
Shunt	$\ll R_g$	$\gg R_g$
Series-shunt {	$\ll R_L$ $\ll r_{off}$	$\gg R_g$ $\gg r_{on}$

15.1.2 Components as electronic switches

The reed switch

The reed switch is a mechanical switch made up of two tongues (or reeds) of nickel-iron encapsulated in a glass tube that is filled with nitrogen or some other inert gas (Figure 15.4). The tongues are normally spaced out but when magnetized by an external magnetic field they will attract each other and make contact. It is usually a current flowing through a small coil surrounding the glass tube that produces the magnetic field.

The main properties of a reed switch are:
- its very low on-resistance (0.1 Ω)
- its very high off-resistance ($>10^9$ Ω)
- the very low offset voltage (<1 μV, mainly due to thermoelectric voltages)
- its high reliability (over 10^7 switching operations).

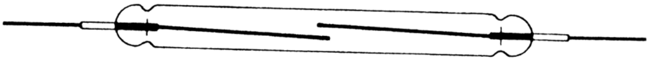


Figure 15.4. The reeds or tongues of a reed switch touch when a magnetic field is introduced.

One disadvantage of this otherwise ideal switch is its low switching speed. A switching frequency of 100 Hz can be achieved but not much more than that. The reed switch is not suitable for high speed switching operations but it is an excellent device for switching functions in, for instance, automatic measurement systems (periodic self calibration, automatic range switching) and telephone switchboards. The reed switch is an inexpensive component, available in a range of encapsulations. There are types that have several switches in a single, IC-like encapsulation incorporating the driving coils.

The photo resistor

The photo resistor (first introduced in Section 7.2) can be used as a switch when combined with a light source – such as an LED – that can be switched. The leading properties of this component as a switch are:
- that it has a rather high on-resistance (up to 10^4 Ω)
- that there is a moderate off-resistance (roughly 10^6 Ω)
- its very low offset voltage (<1 μV).

Due to the inherent slowness of the photoresistive effect (especially when going from light to dark) the switching rate is limited to about 100 Hz.

The PN-diode

It is the high resistance of a pn-diode when reverse biased and its rather low differential resistance when forward biased that makes it suitable for switching operations (see also Section 9.2). Its main properties as a switch are:
- its on resistance which is equal to the differential resistance r_d and inversely proportional to the forward current and 25 Ω at 1 mA

- the high off-resistance of around $10^8\ \Omega$
- that the offset voltage is large, notably equal to the threshold voltage V_k of the forward biased junction which is about 0.6 V.

Another disadvantage is the absence of a separate control terminal. The diode is self-switching, that is to say, it is the voltage across it that makes it switch.

Figure 15.5 shows a switch circuit consisting of four diodes facilitating independent switching (with control currents i_{s1} and i_{s2}) where the offset voltage V_k is compensated. The main typical properties of this switching bridge are:

- the on-resistance: $2r_d//2r_d$, which equals r_d
- the offset voltage: where the difference between the value V_k is usually 1 mV
- the offset current: $i_{s1} - i_{s2}$.

Since diodes have brief delay times, high switching rates are possible (of up to several GHz).

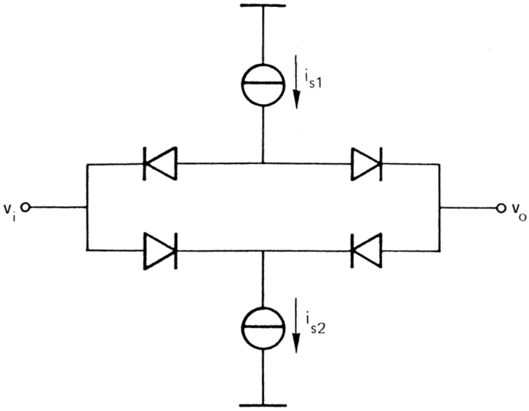


Figure 15.5. The diode bridge is switched off when i_{s1} and i_{s2} are zero. When there is a rather arbitrary positive current the switch will remain on. Its on-resistance is r_d.

The bipolar transistor

To understand how a bipolar transistor can act as a switch we must consider the I_C-V_{CE} characteristic (Figure 15.6a). There are three states: the saturation region, the pinch-off region and the active (or linear) region. When used as a switch, the transistor is either saturated (on) or pinched (off).

Here we shall just consider its use as a shunt switch (Figure 15.6b). In such a case the collector-emitter voltage is $V_{CE} = v_i - I_C R_C$. This is the equation of the so-called load line in Figure 15.6a. It is the base current that controls the switch. The transistor is off for $I_B = 0$, the collector current is also zero (except for the small leakage current) and the transistor is biased at point A. When changing v_i, point A shifts along the pinch-off line. The current does not change and so the device behaves like a high resistance device. For a shunt switch this means $v_o = v_i$.

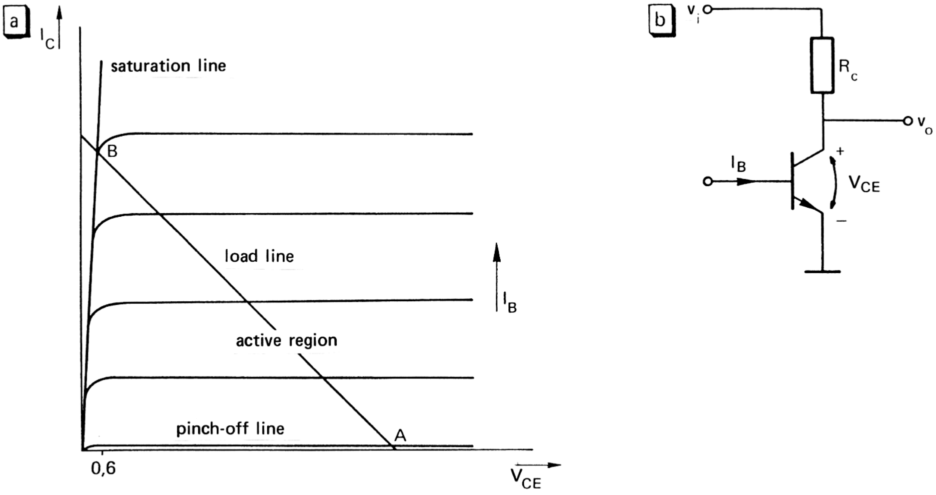


Figure 15.6. The bipolar transistor as a switch, (a) the current-voltage characteristic,
(b) as a shunt switch.

If there is a considerable amount of base current then the collector current will be even greater. The transistor is biased at point B on the saturation line. The slope of the load line is fixed at R_C and that is why point B moves along the saturation line when varying v_i: the resistance is low. V_{CE} hardly changes when the v_i changes and as V_{CE} is almost zero v_o will also be zero.

The bipolar transistor has a high switching speed. Special types have switching rates of several GHz. The leakage currents and the offset voltage make this switch less attractive for high precision applications.

The junction FET

Chapter 11 described junction FET as voltage-controlled resistance, in other words, the channel resistance between the drain and the source depends on the gate voltage (Figure 11.2a). For $V_{GS} = 0$, this resistance is fairly low (50 to 500 Ω). When V_{GS} is below the pinch-off voltage the channel resistance will be almost infinite ($>10^8$ Ω). The ratio between the off and on resistance is therefore high. The advantage that the JFET has over the bipolar transistor is to be found in its very low control current (i.e. in the gate current) and the low offset voltage (generally 1 μV).

Figure 15.7 illustrates how the JFET can be used as a shunt switch and as a series switch. In Figure 15.7a the source voltage is zero, so the gate can simply be controlled by a voltage relative to ground. In Figure 15.7b, the source voltage varies according to the input voltage v_i. To keep the JFET on, V_{GS} must remain zero irrespective of v_i. This can be accomplished by having a resistor between the gate and the source. The switch is on when $V_{GS} = 0$, so for $i_s = 0$. The switch turns off as soon as i_sR exceeds the pinch-off voltage. In this configuration not only the control source but also the signal source must be able to supply the required control current i_s. Here the advantage of powerless control is undermined.

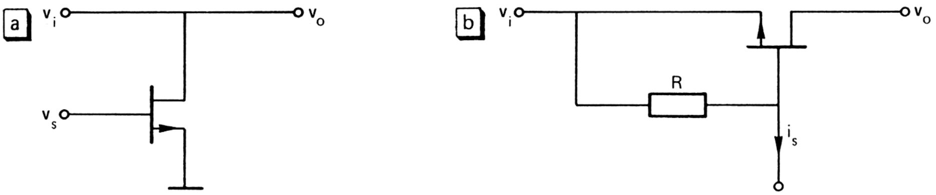


Figure 15.7. A JFET as (a) a shunt switch, (b) a series switch.

The MOSFET

The MOSFET can be employed as a switch in much the same way that the JFET can. With the MOSFET, the control voltage required to switch the device is fairly high and additional substrate voltage (Section 11.1.2) is required. The on-resistance is usually high compared to that of the JFET. The advantages of the MOSFET are, its smaller dimensions and the fact that fewer manufacturing stages are required. This therefore makes it easy to integrate with other components. Finally, its energy consumption is minimal.

MOSFET switches can be found in integrated multiplexers (Section 15.2) and digital integrated circuits (see Chapters 19 and 20). As their on-resistance is considerable, additional buffer stages are required in analog applications so that transfer errors arising during loading can be minimized.

The thyristor

A thyristor can be thought of as a diode which, when forward biased, only conducts after a voltage pulse has been generated at a third terminal (the control gate or the control input). After this pulse has been put out the thyristor will continue to conduct. If the forward current falls below a certain threshold value then the thyristor will switch off and this state will be maintained until there is another control pulse.

A thyristor consists of a four-layer structure made up of alternating layers of p and n-type silicon. The explanation for the physical operation of this device is not given.

Figure 15.8a depicts the circuit symbol for a thyristor (it is like a diode but then with an additional connection). Figure 15.8b illustrates how rectified sine wave switching works.

The thyristor is used particularly in power control for low power systems (like incandescent lamps) and for very high power systems (like electric locomotive engines). The average output power depends on the surface area below the sine waves seen in Figure 15.8b which can be controlled by the phase between the control pulses and the sine wave.

There are similar devices that conduct in two directions (triac) and operate as two thyristors connected in an anti-parallel way.

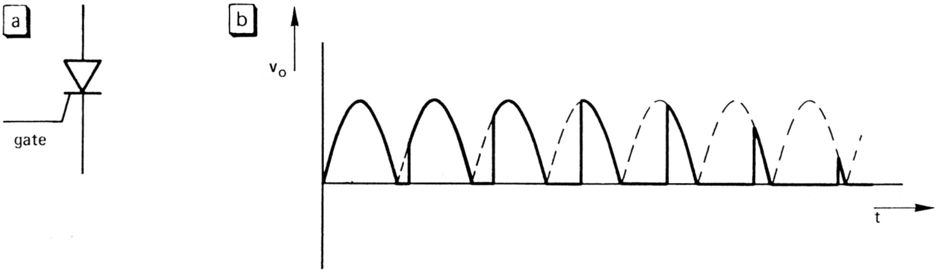


Figure 15.8. (a) The circuit symbol for the thyristor, (b) power control: in this example the mean power is reduced by gradually changing the phase of the control pulse relative to the sine wave.

15.2 Circuits with electronic switches

This section examines two special kinds of circuits in which electronic switches feature as essential components. The first are time-multiplexers (introduced in Section 1.1) and the second are sample-hold circuits. At the end of this section we shall analyze an important phenomenon occurring in most switching circuits, that of the transients derived from capacitive crosstalk.

15.2.1 Time multiplexers

Time multiplexers are used to scan a number of measurement signals and to connect them consecutively to a common information channel so that the expensive parts of an electronic measurement system can be shared (see Section 1.1). The multiplexing (from now on the prefix “time” will be omitted) of digital signals is done with the aid of logic circuits (Chapter 19). Analog multiplexers, available as integrated circuits, require accurate signal transfer for the selected channel.

An analog multiplexer consists of a set of electronic switches that are switched on one at a time and in succession. Figure 15.9a shows a configuration for the multiplexing of measurement signals that constitute the output of a differential amplifier. Figure 15.9b is suitable for the immediate multiplexing of differential voltages (for double-poled switches). It is sometimes also called a differential multiplexer. The switching pairs of a differential multiplexer must have equal on and off-resistances so that they do not degenerate the CMRR of the circuit. Some types of multiplexers allow the user to choose between single or double mode multiplexing.

An integrated multiplexer contains various additional circuits such as, a decoder to select the correct switch and switch drivers to supply the voltages required to the electronic switches (Figure 15.10). Usually the channel selection is controlled by a computer that sends a binary coded signal to the decoder input. With p binary lines 2^p different codes can be transmitted. A multiplexer with n individual channels only needs $2\log n$ control inputs to select 1 channel out of n. The decoder thus substantially reduces the number of multiplexer IC connection pins.

Many multiplexers have additional control input known as the "enable input". With the binary control signal all channels can be switched off simultaneously, irrespective of channel selection. This feature facilitates the multiplexing of a number of multiplexer ICs and thus also, an extension of the number of channels.

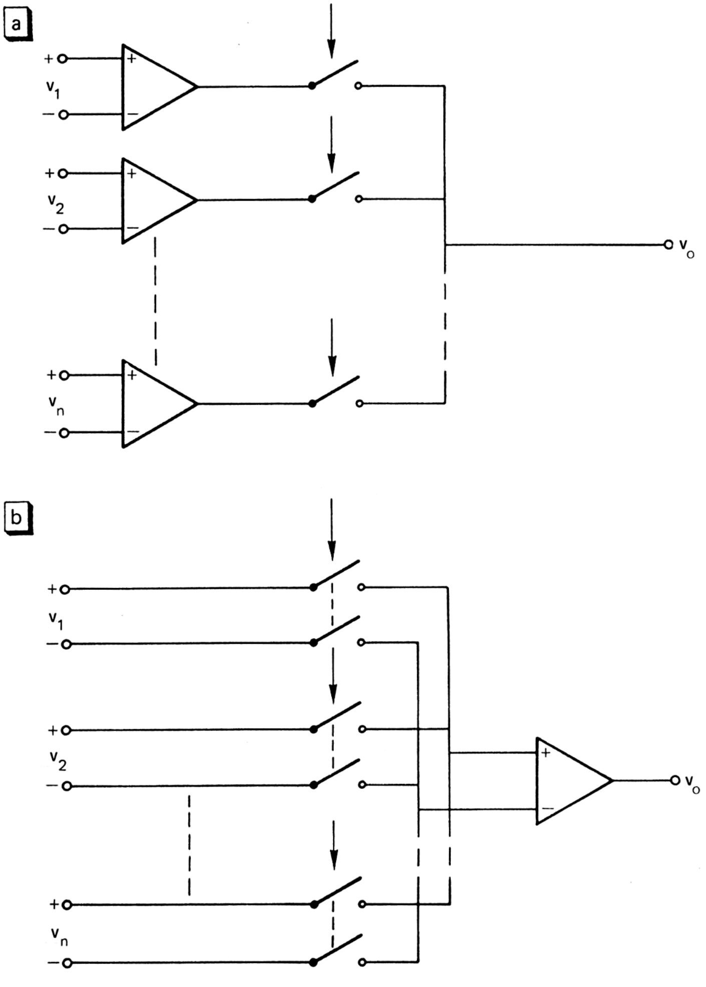


Figure 15.9. A multiplexer for *n* channels (a) with differential input amplifiers and single-poled switches, (b) with double-poled switches and a single differential amplifier.

Table 15.2 displays the properties of a multiplexer circuit where each switch is composed of an n-channel and a p-channel MOSFET in parallel. Most analog multiplexers can be used as demultiplexers as well (see, for instance, the circuit in Figure 15.10).

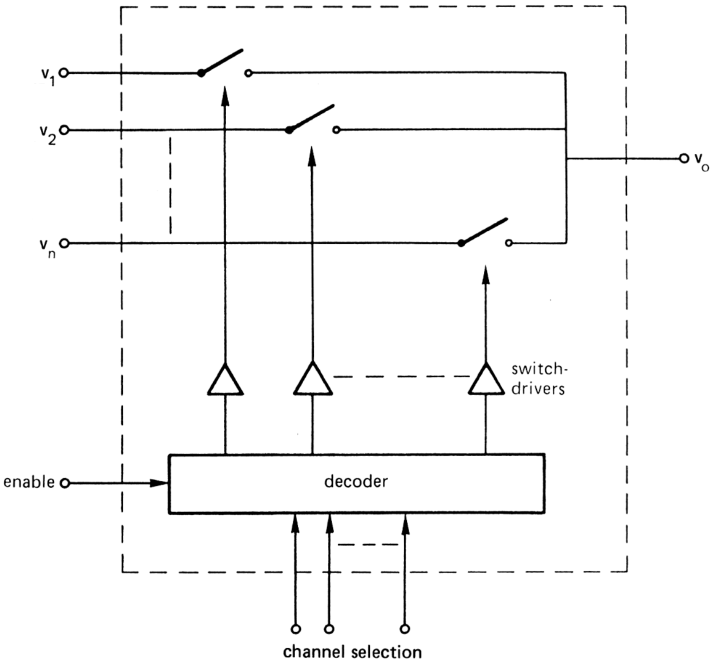


Figure 15.10. The simplified internal structure of an integrated multiplexer.

Table 15.2. The specifications for an integrated 16-channel multiplexer.

number of channels	16
Contacts	break before making
voltage range	–15 to + 15 volts
r_{on}	700 Ω + 4%/K
Δr_{on}	< 10 Ω
I_{off} (input and output)	0.5 nA
CMRR (DC)	125 dB
(60Hz)	75 dB
$C_i(off)$	2.5 pF
$C_o(off)$	18 pF
$C_{transfer}(off)$	0.02 pF
Setting time (0.01%)	800 ns
(0.1%)	250 ns
power dissipation	525 mW

15.2.2 Sample-hold circuits

A sample-hold circuit is a signal processing device with two different transfer modes (i.e. with two states). In the sample/track mode, the output follows the input and usually the transfer is 1. In the hold mode the output retains the value at the moment of hold command. Figure 15.11 illustrates how a sample-hold circuit operates.

When looking for the possible errors in such a circuit, four phases can be distinguished: the track phase, track-hold transition, the hold-phase and hold-track transition (Figure 15.12).

(1) the track phase: in this state similar errors occur to those seen with an operational amplifier, errors like offset voltage, drift, bias currents, noise, gain errors and limited bandwidth;
(2) track-hold transition: the main errors are the delay time t_1 (also aperture delay time) and the uncertainty attached to that delay time, the aperture jitter t_2;
(3) the hold phase: during this phase the hold voltage can drift away as denoted by the term droop;
(4) hold-track transition: the specified times for this are given in Figure 15.12.

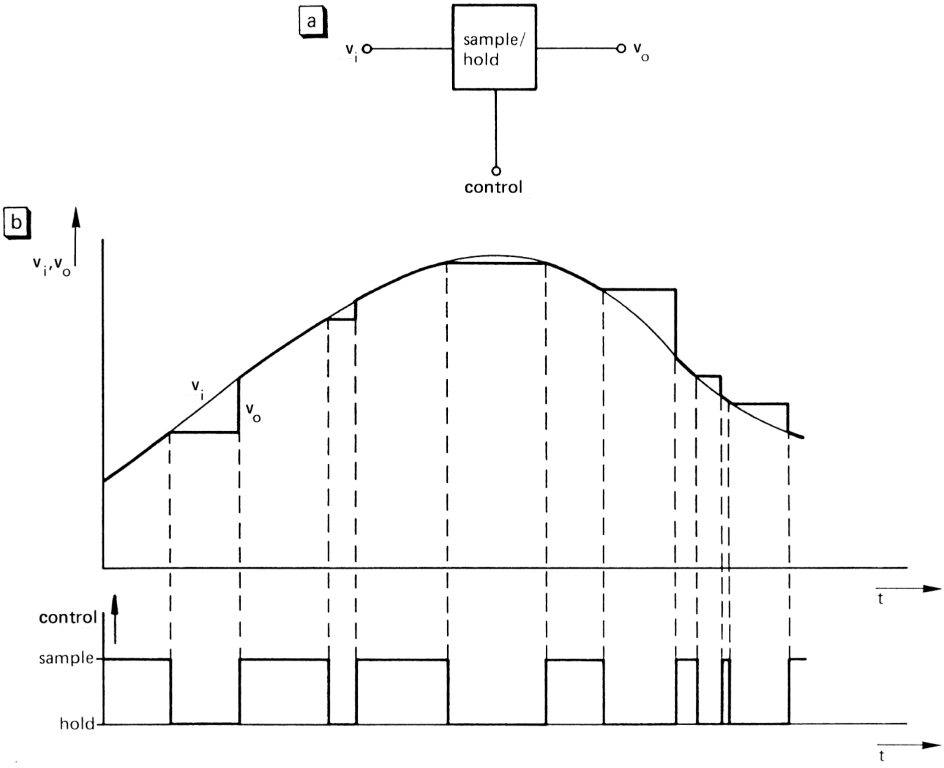


Figure 15.11. (a) The circuit symbol for a sample-hold circuit, (b) an example of the output voltage v_o where there is a corresponding input voltage v_i and a control signal.

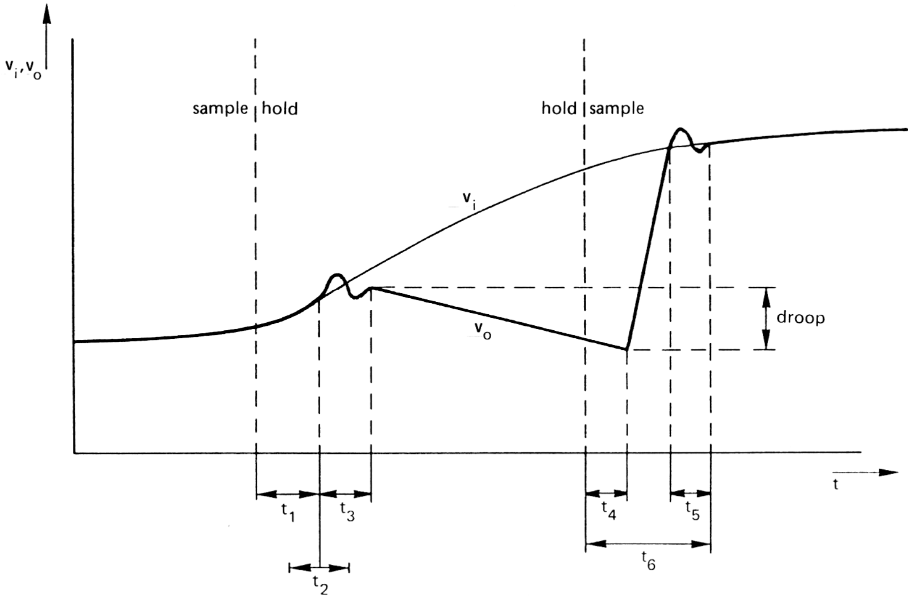


Figure 15.12. The various transition times and delay times of a sample-hold circuit: t_1 = turn-off delay time or aperture delay time, t_2 = aperture uncertainty or aperture jitter, t_3 = settling time, t_4 = turn-on delay time, t_5 = settling time, t_6 = acquisition time.

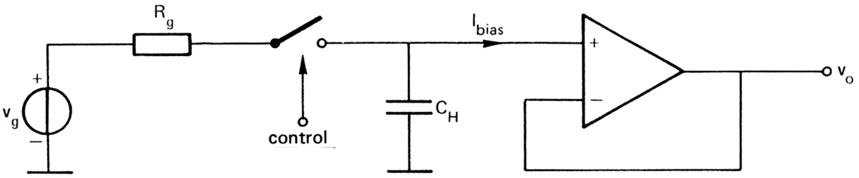


Figure 15.13. A simple sample-hold circuit with a capacitor as a memory device.

In Figure 15.13 we see a simple version of a sample-hold circuit that is composed of a switch, a hold capacitor C_H and a buffer amplifier. The capacitor acts as the analog memory for the voltage to be retained. It is charged by the input source via the switch. The time constant of the charging (Section 4.2) is $(R_g + r_{on})C_H$; the smaller the C_H is, the faster the capacitor will be charged to the input voltage.

When the switch is off the capacitor will remain charged because both the switch and the amplifier have high resistances. However, the bias current of the buffer amplifier tends to discharge the capacitor (or charge it, depending on the direction of the bias current). During the hold period T_H, the voltage of the capacitor changes by an amount which is $\Delta v_o = T_H I_{bias}/C_H$. Therefore, if there is to be a small droop at given I_{bias}, the capacitance has to be as large as possible.

Table 15.3 gives an overview of some of the specifications belonging to a particular type of sample-hold device with a structure such as that given in Figure 15.14.

Table 15.3. Some integrated sample-hold circuit specifications based on the structure given in Figure 15.14.

Analog input	
V_{off}	6 mV
I_{bias}	3 μA
R_{i}	30 MΩ
hold → sample	
t_{acq} (0.01%)	25 μs
(0.1 %)	6 μs
sample → hold	
aperture delay	150 μs
aperture jitter	15 μs
settling time (0.01%)	0.5 μs
hold mode:	
I (droop)	100 pA
track-mode	
CMRR	60 dB
Bandwidth	1.5 MHz
Slew rate	3V/μs
R_{o}	12 Ω

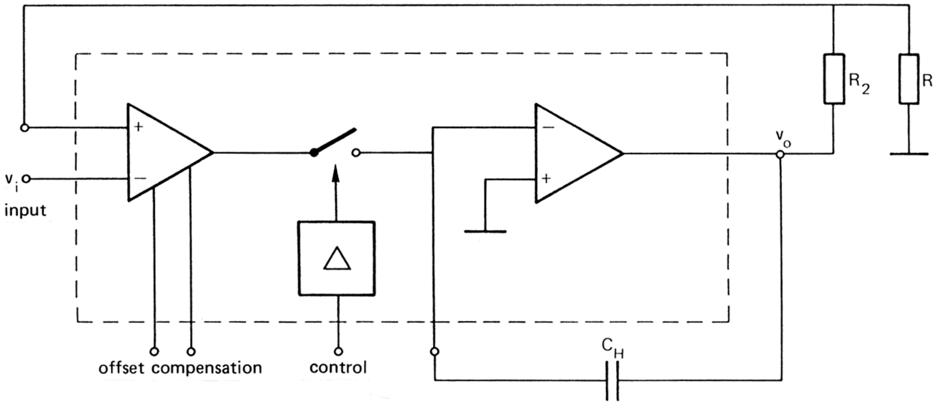


Figure 15.14. An integrated sample-hold circuit. All components are integrated except for the hold capacitor and some resistors that have to be connected externally.

Switching to the hold mode requires a control input voltage of more than 2 V. To change to the track mode the circuit requires a control voltage of below 0.8 V. The input is provided with a integrated input amplifier. The user can set the transfer in the track mode to an arbitrary value by connecting the proper values of resistors R_1 and R_2. The capacitor C_H should also be externally connected.

15.2.3 Transient errors

Due to the capacitive crosstalk emanating from the rectangular-shaped control signals of switches, electronic switches often produce unwanted pulse-shaped voltages that are superimposed on the measurement signal. Such crosstalk in a series switch, controlled by a voltage v_s is illustrated in Figure 15.15.

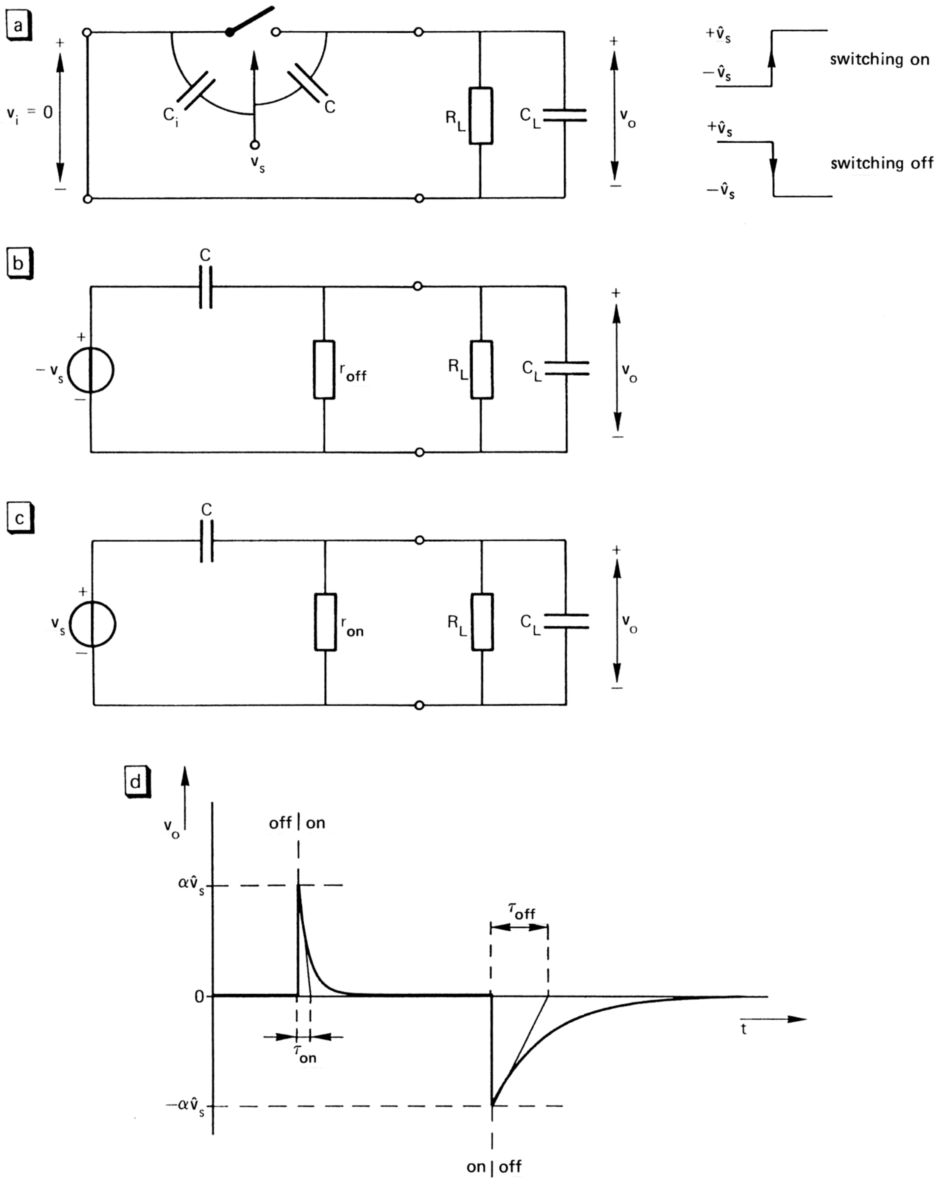

a
+
$v_i = 0$
−
C_i
C
v_s
R_L
C_L
+
v_o
−
$+\hat{v}_s$
$-\hat{v}_s$
switching on
$+\hat{v}_s$
$-\hat{v}_s$
switching off
b
C
+
$-v_s$
−
r_{off}
R_L
C_L
+
v_o
−
c
C
+
v_s
−
r_{on}
R_L
C_L
+
v_o
−
d
v_o
off | on
$\alpha\hat{v}_s$
τ_{off}
0
t
τ_{on}
$-\alpha\hat{v}_s$
on | off

Figure 15.15. Transients derived from switching, (a) to account for the crosstalk of v_s to the output where the input is short-circuited, (b) a circuit model directly after having been switched off, (c) a circuit model directly after having been switched on, (d) an output voltage derived from capacitive crosstalk.

We shall calculate the output signal derived just from the control signal. The input voltage is set at zero – or the equivalent – and the input terminal is ground connected. Figures 15.15b and c show the situations directly after the respective off and on-commands have been given. It is presumed that the switch itself responds instantly. R_L and C_L represent the switch load (or, the input impedance of the connected circuit). In both situations, the circuit corresponds to a differentiating network as given in Figure 8.7. The output voltage due to a step in the control voltage is:

$$v_o = \alpha v_s e^{-t/\tau} \tag{15.7}$$

with $\alpha = C/(C + C_L)$.
After switching off, $\tau = \tau_{off} \approx R_L(C + C_L)$; after switching on, $\tau = \tau_{on} \approx r_{on}(C + C_L)$. The step responses are depicted in Figure 15.15d.
For $C_L = 0$ (no capacitive load), the level of the output pulse is equal to the level of the control voltage (which can be several volts). A capacitive load somewhat reduces the crosstalk pulses. The resulting sharp pulses can be capacitively coupled to other parts of the circuit, so it is important to keep them as low as possible.

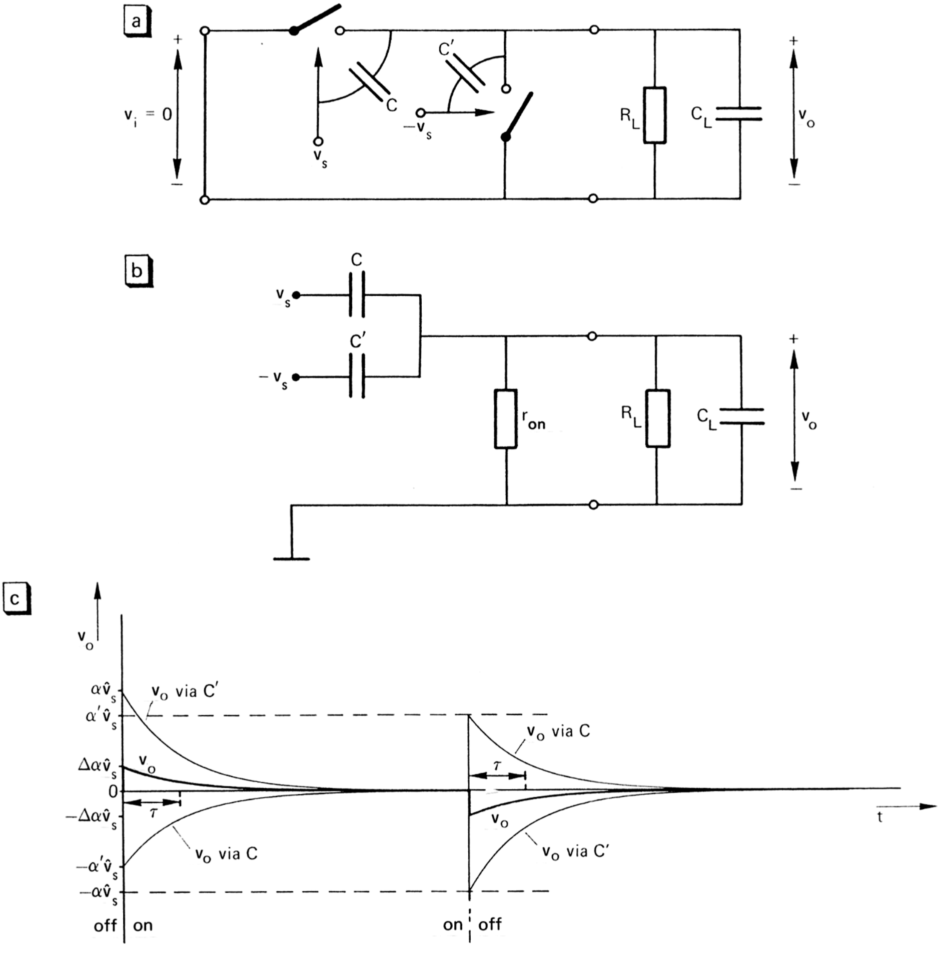

a
+
v_i = 0
−
C
C′
v_s
−v_s
R_L
C_L
+
v_o
−
b
C
v_s
C′
−v_s
r_on
R_L
C_L
+
v_o
−
c
v_o
αv̂_s
α′v̂_s
Δαv̂_s
0
−Δαv̂_s
−α′v̂_s
−αv̂_s
v_o via C′
v_o
τ
v_o via C
v_o via C
τ
v_o
v_o via C′
t
off
on
on
off

Figure 15.16. A series-shunt switch with compensation transients, (a) a switch with short-circuited input, (b) a model for both the on and off states, (c) an output signal derived from control voltages v_s.

In the series-shunt switch to be seen in Figure 15.16 both switches produce crosstalk pulses but they do compensate each other.

The circuit model in Figure 15.16a only contains the relevant components and Figure 15.16b shows the model used for the analysis. This model is valid for both the on and off-states. In either state, one switch is off while the other remains on and the switch resistances are in parallel. The output voltage is obtained by adding the contributions made by the two signal voltages:

$$v_o = \alpha v_s e^{-t/\tau} - \alpha' v_s e^{-t/\tau} \tag{15.8}$$

with $\alpha = C/(C + C' + C_L)$, $\alpha' = C'/(C + C' + C_L)$ and $\tau = \tau' = r_{on}(C + C' + C_L)$. Here it is assumed that $r_{on} << R_L$. The compensation is perfect if C and C' are equal, in that case $\alpha = \alpha'$, hence $\Delta\alpha$ in Figure 15.16 is zero. This may only be said to apply if both switches respond exactly simultaneously and have equal on-resistances.

SUMMARY

Electronic switches

- The important parameters of electronically controlled switches are the on and off-resistances, the offset voltage and the leakage current. In the case of dynamic behavior, the delay times and the acquisition time constitute important characteristics.
- When combined with a source resistance and a load resistance the on and off-resistances create transfer errors.
- The following components can be employed as electronic switches: the reed switch, the photo-resistor, the pn-diode, the bipolar transistor, the JFET and MOSFET, and the thyristor.
- The reed switch and the photo-resistor are slow switches, transistors are faster and diodes have the highest speeds of all.
- When used as a switch, the gate-source voltage of the JFET acts as the control quantity: the JFET in fact switches between low channel resistance (50 to 500 Ω) and very high values.
- The MOSFET is widely used as an electronic switch, notably in integrated circuits where there are many active components, like in microprocessors.
- The thyristor can be thought of as a diode that only conducts after a pulse has been sent to the control gate. It switches itself off when the voltage or current falls below a certain value. Thyristors are used for power control where they control the moment of ignition with respect to the power signal phase.

Circuits with electronic switches

- A time multiplexer is a multiple switch with multiple inputs and a single output. Channel selection is performed using a decoder.
- A sample-hold circuit has two states (or modes) known as track and hold mode. The output tracks the input during the track mode; this value is retained at output

when the hold mode is switched to.

- Sample-hold circuits are available as complete integrated circuits. The important parameters are their properties as followers (or voltage amplifiers), the delay and acquisition times and droop (i.e. the drift during the hold mode).

EXERCISES

Electronic switches

15.1 You are given a voltage source with source resistance 10 Ω. This voltage must be connected to a circuit with an input resistance of 50 kΩ that uses an electronic switch in series with the source and load. Determine the conditions that must exist for the on and off-resistance of the switch if the following requirements are to be met. The maximum transfer error in the on-state is 0.1% and there is a maximum transfer of 0.1% in the off-state.

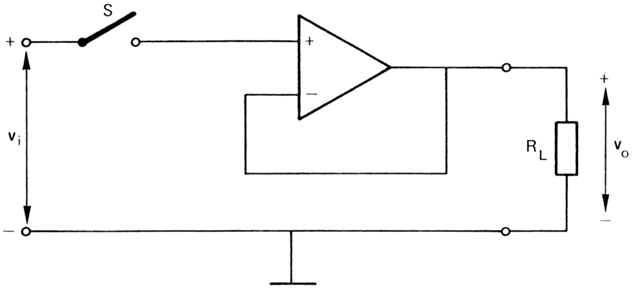
15.2 Answer the same question but this time for a shunt switch.

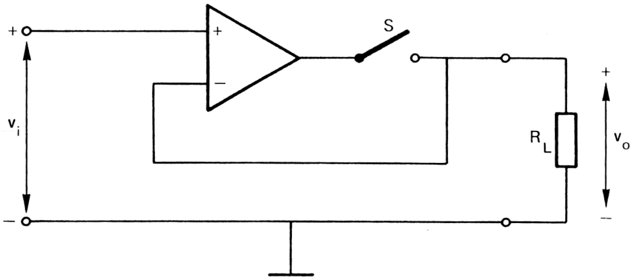
15.3 Answer the same question but this time for a series-shunt switch consisting of two identical switches (Figure 15.3c).

15.4 Create a table that is similar to Table 15.1 but this time for a current switch. The current source (parallel) resistance is R_g and the load resistance is R_L; $R_L << R_g$).

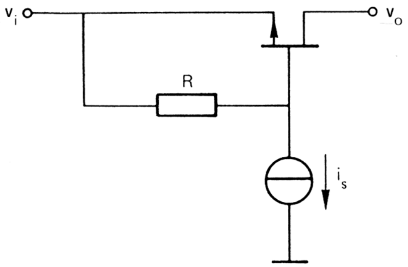
15.5 What is the on-resistance of the diode bridge when there is a control current of 5 mA?

15.6 The following figures give two possible combinations for an electronic switch and an operational amplifier. Discuss the effect that the on-resistance and the offset voltage of the switch have on the accuracy of the transfer in the on-state. Which configuration is preferable, and why?

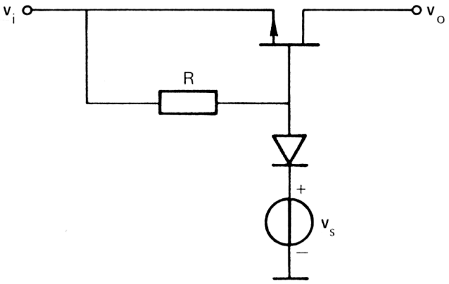




15.7 The pinch-off voltage of the JFET used in the circuit below has a specific range of between −2 and −6 V. The control current of this switch is i_s = 2 mA. Find the proper value of *R* to guarantee correct switching.

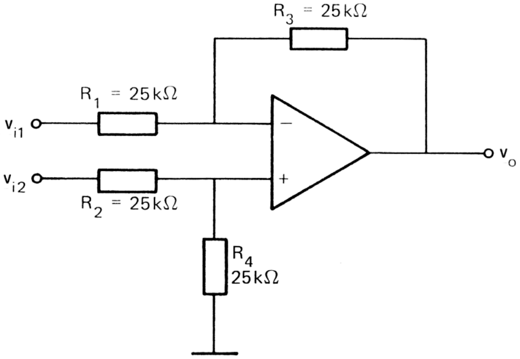


15.8 The control quantity for the JFET switch referred to in Exercise 15.7 is a voltage connected in the way shown in the figure below. The input voltage varies between −3 and +3 V. Find the conditions for the upper and lower levels of the control voltage if the switch is to operate properly.



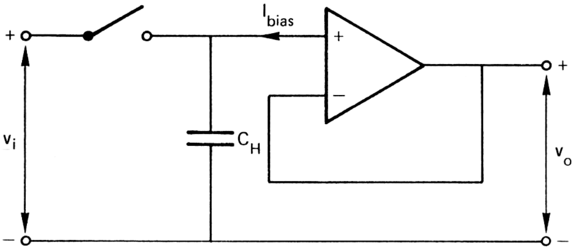
Circuits with electronic switches

15.9 The output voltage of a double-poled multiplexer is connected to the differential amplifier in the way given in the figure below. The on-resistance of the multiplexer is specified as r_{on} = 500 Ω ± 5 Ω. The operational amplifier is ideal. Calculate the CMRR for the whole circuit.



15.10 The specifications given in the following sample-hold circuit are: V_{off} = 100 µV; I_{bias} = 1 µA (positive in the direction of the arrow); C_H = 1 µF and r_{on} =

100 Ω. The circuit is connected to a voltage source with a source resistance of 50 Ω. Calculate the absolute error in the output voltage when the circuit is in the track mode.



15.11 Imagine that the circuit referred to in Exercise 15.10 is now connected to a voltage source with a sinusoidal output where the amplitude is 8 V and the frequency is 100 Hz. The offset voltage is adjusted to zero. Just at the moment when v_i reaches its peak value, a hold command is given. Find v_o after 1 and after 100 input signal periods.

16 Signal generation

Measurement signals are almost invariably aperiodic but pure periodic signals are also important in instrumentation. More often than not, they serve as auxiliary signals like, for instance, when functioning as modulated signal carriers (see in this connection Chapter 17) or when used as test signals, for example to analyze a system's frequency transfer function (see for further details Chapter 21). It is therefore worth knowing how periodic signals are generated. An instrument that produces a periodic signal is called a signal generator. If the signals produced are just sinusoidal then such an instrument is known as an oscillator. Indeed, the principle of precisely how oscillators work will be explained in the first part of this chapter. Finally, there are instruments that are able to generate other periodic signals. Function generators, to name but one sort, produce divergent periodic signals such as: square wave, triangular, ramp and sine-shaped signals. All these kinds of instruments will be described in the second part of this chapter.

16.1 Sine wave oscillators

There are various ways to generate a sinusoidal signal, one way is by solving a second order differential equation using analog electronic circuits, a principle that is outlined in this section. A second way involves starting with a symmetric, rectangular or triangular signal. The sine shape is obtained either by filtering out the superharmonics and keeping the fundamental, or by reshaping the signal using resistance-diode networks of the type described in Chapter 14. Obviously, the accuracy of this last method will depend very much on the quality of such non-linear converters. A third way is by synthesizing arbitrary periodic signals with the aid of a computer. There the processor generates a series of successive codes that are converted into an analog signal by a DA converter (Chapter 18).

16.1.1 Harmonic oscillators

The general solution to the linear differential equation

$$a_0 \frac{d^2x}{dt^2} + a_1 \frac{dx}{dt} + a_2 x = 0 \tag{16.1}$$

is

$$x(t) = \hat{x} e^{-\alpha t} \sin(\omega t + \varphi) \tag{16.2}$$

in which $\alpha = a_1/2a_0$, $\omega = \sqrt{}(a_2/a_0 - a_1^2/4a_0^2)$ and $\hat{x}$ and φ are arbitrary constants. In accordance with the sign α, $x(t)$ remains a sinusoidal signal with an exponentially decreasing amplitude ($\alpha > 0$) or an exponentially increasing amplitude ($\alpha < 0$). Only when $\alpha = 0$, $x(t)$ is a pure sine wave with constant amplitude. Consequently, the coefficient α is termed the damping factor. It is fairly easy to design an electronic circuit with voltages and currents to satisfy equation (16.1). In order to generate pure sine waves at a constant amplitude the coefficient a_1 should be kept at zero which does require extra effort.

The signal derivatives and integrals are obtained from inductances and capacitances. Active elements are required to keep the damping factor at zero. In addition to this, some kind of feedback appears to be necessary. An example will be given to illustrate the basic principle (in practical terms the example is of little significance but it is useful for explanatory purposes). Figure 16.1 provides a block diagram of an electronic system with two differentiators and one amplifier in series.

The output is connected straight to the input which means that:

$$v_o = K\tau^2 \frac{d^2 v_o}{dt^2} \tag{16.3}$$

The solution to this linear, homogeneous differential equation is $v_o = \hat{v} \sin(\omega t + \varphi)$, with $\omega^2 = -1/K\tau^2$. Evidently K must be negative (it is an inverting amplifier). For $K = -1$ the frequency of the signal produced is $f = 1/2\pi\tau$. Instead of using differentiators one can alternatively take integrators. As was seen in Chapter 13, an integrator is more stable and produces less noise than a differentiator. The differential equation, however, remains the same.

The amplitude can have any value between the system's signal limits, it is not fixed by circuit parameters. As long as a_1 in Equation (16.1) is zero the amplitude, once present, will remain constant. In an actual circuit, though, the component values will vary perpetually, for instance because of temperature fluctuations, so the condition $a_1 = 0$ will not be met for a long time. This means that somehow the system has to be controlled if the amplitude is to be fixed at a prescribed value. There is also another reason why such a control circuit is needed. When the system is switched on the amplitude is zero. As long as the damping factor α is zero, or positive, the amplitude will remain at zero. Therefore α must be negative for a short period so that the amplitude can be allowed to rise to the desired value. When that value has been reached α must revert again to zero.

In the circuit given in Figure 16.1 the term a_1 is obtained by simply adding a fraction β of v_2 to the voltage v_3. The output voltage thus becomes:

$$v_o = K\left(\tau^2 \frac{d^2 v_o}{dt^2} + \beta\tau \frac{dv_o}{dt}\right) \tag{16.4}$$

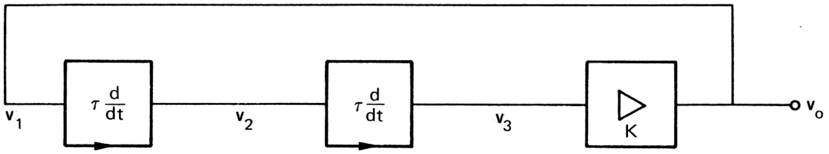


Figure 16.1. An oscillator with two differentiators and an amplifier.

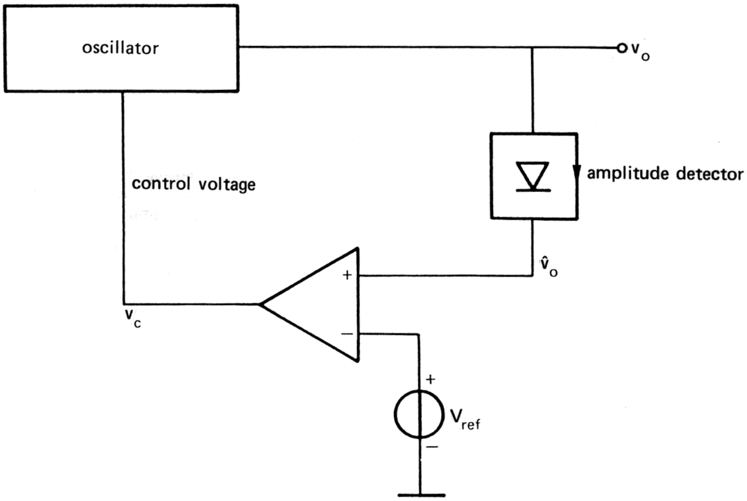


Figure 16.2. The principle of an oscillator with amplitude control.

Whether the output amplitude increases or decreases depends on the fraction β. With $\beta = 0$ a sine wave with steady output is generated.

Figure 16.2 shows how an oscillator with electronic amplitude control is set up. The control system consists of the following components:

- the amplitude detector. Its output $\hat{v}_o$ is a measure of the amplitude of the generated sine wave (like, for instance, the peak detector shown in Section 9.2.2 or a rectifier with low-pass filter);
- a reference voltage V_{ref};
- a control amplifier to amplify the difference between the reference voltage V_{ref} and the peak value $\hat{v}_o$;
- a control element in the oscillator, this can be an electronically controllable resistance (JFET, thermistor; photoresistor) or an analogue multiplier. This control element affects the factor a_1 in Equation (16.1) or β in Equation (16.4) and therefore also the damping factor α.

As the signal power is determined by the square of the signal voltage amplitude or the current amplitude a control element based on heat dissipation may be utilized. An element that is widely used for this purpose is the thermistor which is part of the oscillator network that operates in such a way that at increasing amplitude (i.e. at decreasing resistance) further increase is halted. This method, requiring merely a single component, is extremely simple but control is slow due to the thermal nature of the network. What should also be pointed out is that in a steady state the amplitude depends on the thermistor parameters as well as on the heat resistance to the environment (i.e. the environmental temperature). This method is not suited to high amplitude stability.

16.1.2 Harmonic oscillator circuits

The relationships between the voltages and the currents in an electronic network are given as linear differential equations. We have introduced complex variables (see Chapter 4), mainly because the solving of these equations inevitably becomes rather time-consuming. In its steady state, a harmonic oscillator generates a pure sine wave which means that oscillators of this kind can be analyzed with the help of complex variables. Equation (16.3), for example, which belongs to the circuit given in Figure 16.1, can be written as $V_o = K\tau^2(j\omega)^2V_o$ which therefore means that $\omega^2 = -1/K\tau^2$.

Any harmonic oscillator is composed of at least one amplifier and one passive network with a frequency-selective transfer. In most cases, an oscillator can be modeled in the way shown in Figure 16.3. When mutual loading can be ignored or when this effect can be discounted by A or β then $V_o = AV_i$ and $V_i = \beta(\omega)V_o$ so that $A\beta(\omega) = 1$. This complex equation is termed the oscillation condition. The conditions for oscillation and oscillation frequency are provided when this equation is solved. Several examples will now be given to illustrate this point.

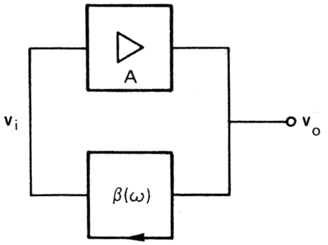


Figure 16.3. A basic harmonic oscillator diagram showing how the amplifier output is fed back to the input via a network by means of frequency selective transfer.

The Wien oscillator

With a Wien oscillator, the feedback comes from two resistors and two capacitors arranged as voltage dividers with band-pass characteristics (Figure 16.4). Under the conditions $R_1 = R_2 = R$ and $C_1 = C_2 = C$, the transfer function of this Wien network is

$$\beta(\omega) = \frac{V_i}{V_o} = \frac{1}{3 + j\omega\tau + 1/j\omega\tau} \tag{16.5}$$

The oscillation condition is $A\beta(\omega) = 1$, hence:
$3 + j\omega\tau + 1/j\omega\tau = A$
The real parts on the left and right-hand sides of this equation must be equal, just like the imaginary parts. This will result in two equations:

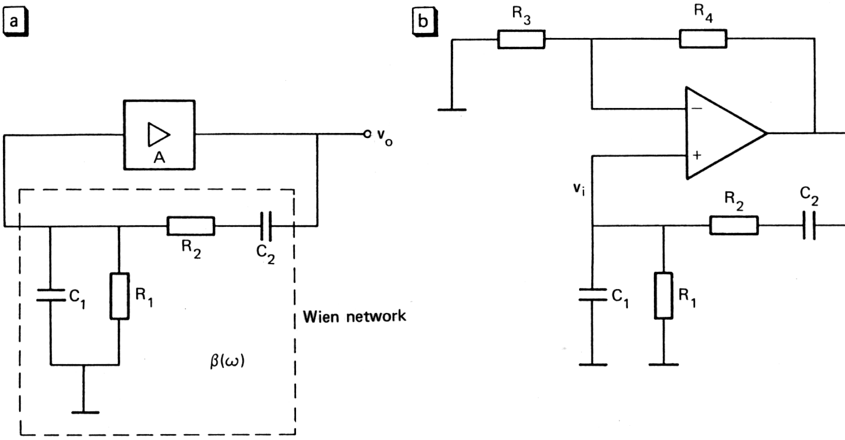


Figure 16.4. (a) An oscillator created according to the principle given in Figure 16.3 with a Wien network, (b) a Wien oscillator that has one operational amplifier.

$$A = 3 \tag{16.6}$$

$$\omega = \frac{1}{\tau} \tag{16.7}$$

Equation (16.6) describes the condition for constant amplitude. If $A > 3$ the amplitude will increase while with $A < 3$ it will decrease. Equation (16.7) gives the frequency of the signal generated. Figure 16.4b shows a possible Wien oscillator configuration without amplitude control. The calculation of the oscillation conditions arising directly from this circuit results in the equations $\omega = 1/\tau$ and $R_4 = 2R_3$. This is not a surprising result, without the Wien network the amplifier acts as a non-inverting amplifier with a gain of $1 + R_4/R_3 = 3$ for the condition previously found.

The phase-shift oscillator

The basic idea underlying the phase-shift oscillator is depicted in Figure 16.5. Here the feedback network consists of three cascaded low-pass RC filters. If the values of the resistors and capacitors are such that the time constants are equal but the sections do not load each other (cf. Figure 8.13) then the transfer will be $\beta(\omega) = 1/(1 + j\omega\tau)^3$. The oscillation condition is simply $K = (1 + j\omega\tau)^3$. If this equation is divided into real and imaginary parts then this will result in $K = -8$ and $\omega^2\tau^2 = 3$.

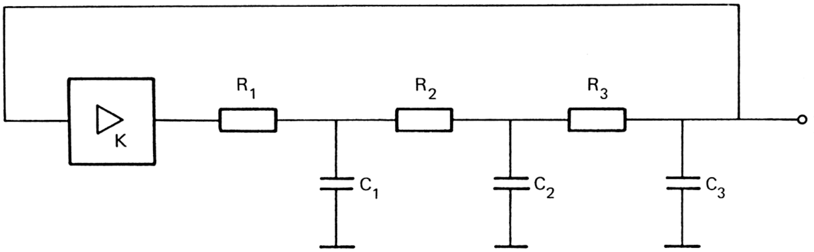


Figure 16.5. A phase-shift oscillator with three low-pass RC-sections.

The two-integrator oscillator

The last example of a harmonic oscillator to be given is the two-integrator oscillator or the dual integrator loop (Figure 16.6). The oscillator condition can easily be established: $(-1/j\omega\tau)^2(-1) = 1$, hence $\omega = 1/\tau$. The oscillation frequency can be varied by simultaneous changing the resistors R_1 and R_2 or the capacitors C_1 and C_2. The same effect can be achieved by having two adjustable voltage dividers (potentiometers) situated at the integrator inputs (Figure 16.7). If k is the attenuation of the voltage dividers then the oscillation condition will be $(1/j\omega\tau)^2k^2(-1) = 1$ so that $\omega = k/\tau$. The oscillation frequency is proportional to the attenuation of the potentiometers.

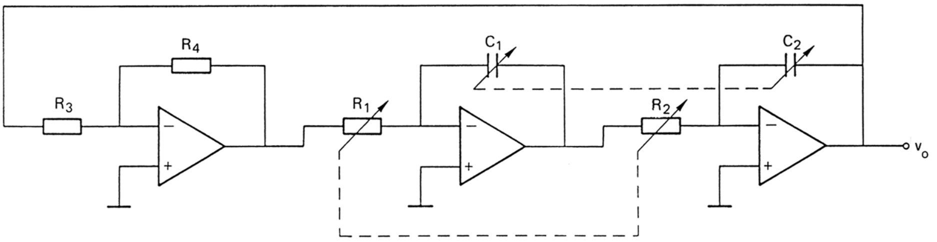


Figure 16.6. A two-integrator oscillator consisting of an amplifier and two integrators. The frequency is adjustable with both the capacitors *C* and the resistors *R*. Usually C_1 and C_2 are switched in stages by a factor of 10 (rough frequency adjustment), whereas R_1 and R_2 remain the potentiometers for the fine tuning of the frequency.

In Figure 16.7 an amplitude stabilization circuit is also given. A fraction β of v_2 is added via R_5 to the leftmost inversion (cf. Equation 16.4). The value of β is multiplied by the output of the control amplifier where the inputs are the rectified oscillator output voltage and a reference voltage. With the control circuit the difference between V_{ref} and $\hat{v}_o$ tends to be zero.

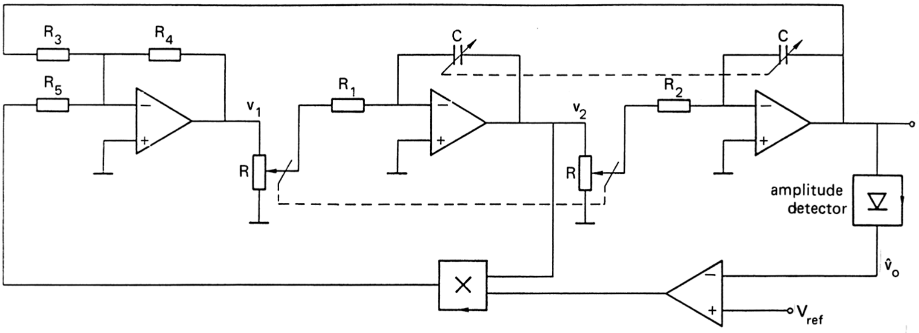


Figure 16.7. A possible configuration for the two-integrator oscillator with an amplitude stabilization circuit.

16.2 Voltage generators

Periodic non-sinusoidal signals are frequently used in instrumentation systems. In combination with an oscilloscope (see Chapter 21) it is possible to visualize a system's step and pulse responses or to measure its rise and delay time. This is done by connecting a periodic square wave or pulse signal to the test system input and observing its output on an oscilloscope or a computer monitor. Triangular or ramp signals make it possible to determine a system's non-linearity. They can also function as control signals in various actuators and be used for the testing of systems or products. Pulse and square wave signals are widely used in digital systems, for instance for synchronization.

In this section we shall discuss a number of the generators used with non-sinusoidal periodic voltages. Most of these instruments are based on the periodic charging and discharging of a capacitor.

16.2.1 Triangle voltage generators

A triangle generator periodically charges and discharges a capacitor with a constant current (Figure 16.8). At constant current, the voltage across a capacitance constitutes a linear time function. This voltage is connected to a Schmitt-trigger (14.2.2) with output levels V^+ and V^- (Figure 16.9). The control circuit controls the switches in such a way that for $v_s = V^+$ the capacitor is charged with current I_1 and for $v_s = V^-$ it is discharged with current I_2. This results in a triangular voltage across C. Any time v_o passes the Schmitt-trigger switching levels the two states will be automatically interchanged. The peak values of the triangle are equal to the switching levels of the Schmitt-trigger. The slope of the triangle can be adjusted by varying the values of currents I_1 and I_2.

Figure 16.10 shows a simple configuration with one integrator, a Schmitt-trigger and an inverting amplifier. At equal charging and discharging currents and $V^+ = V^-$, a symmetrical triangle voltage is obtained. The circuit simultaneously generates a square wave voltage.

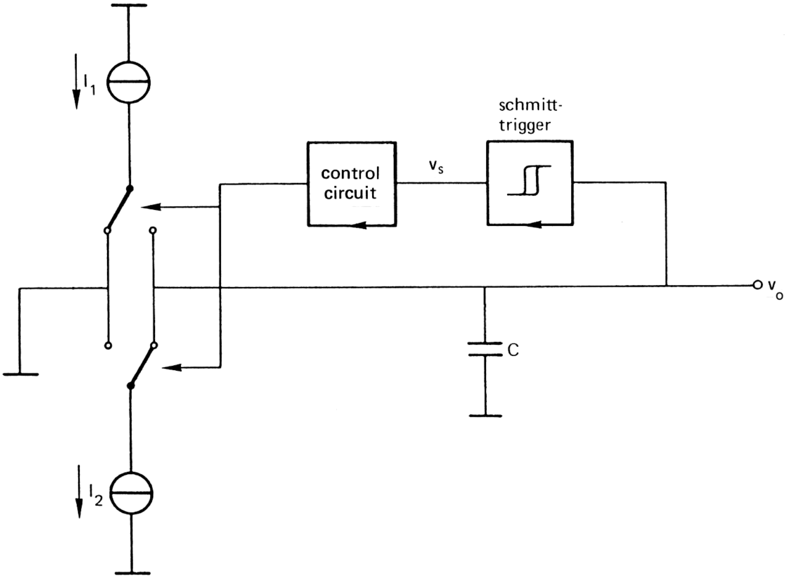


Figure 16.8. A functional diagram of a triangle generator. The triangular voltage is produced by periodically charging and discharging a capacitor with a constant current.

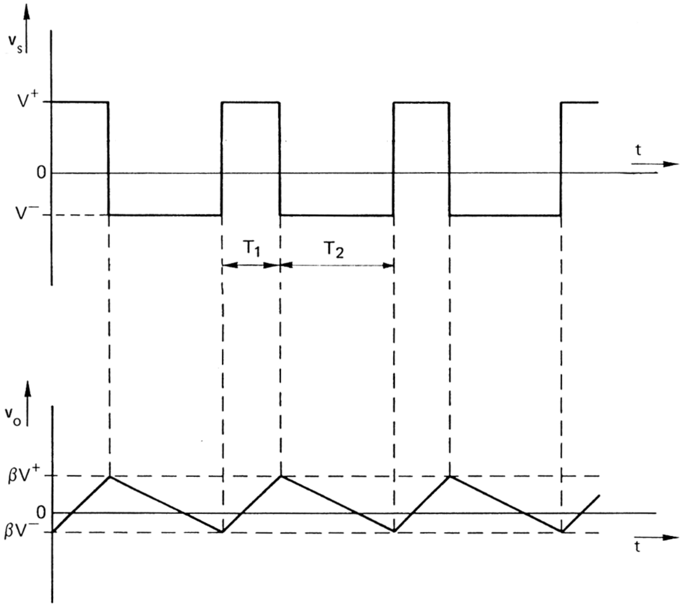


Figure 16.9. The voltages seen in Figure 16.8. V^+ and V^- are the positive and negative power supply voltages, $ßV^+$ and $ßV^-$ are the Schmitt-trigger switching levels. The rise and fall time of v_o is determined by I_1, I_2 and C.

The ratio between the time T_1 and the total period time $T = T_1 + T_2$ of such periodic signals is the signal's duty cycle. A symmetric signal has a duty cycle equivalent to 50%. The duty cycle can be changed by varying one of the currents in Figure 16.8.

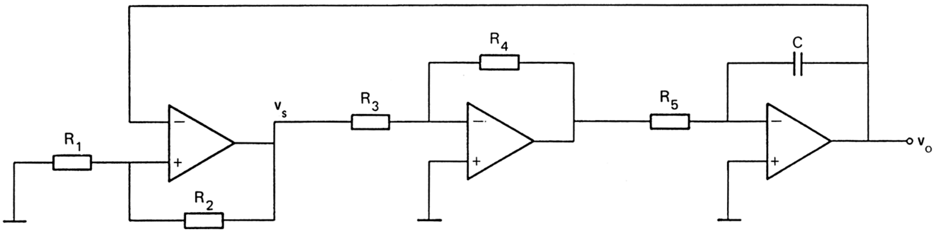


Figure 16.10. A simple configuration for a triangle generator using a Schmitt-trigger, an inverting amplifier and an integrator.

16.2.2 The ramp generator

A ramp voltage can be viewed as a triangular voltage with one vertical slope. Such a short fall time is obtained by discharging a charged capacitor over a switch. The switch is controlled by a Schmitt-trigger (Figure 16.11). For $v_s = V^+$ the switch is on and for $v_s = V^-$ it is off. The capacitor is part of an integrator that has an input connected to a fixed reference voltage V_{ref1}. The output of the integrator rises linearly over the course of time until the switch goes on and the capacitor discharges. In that way, the output reverts to zero. The process will start all over again when the switch is released.

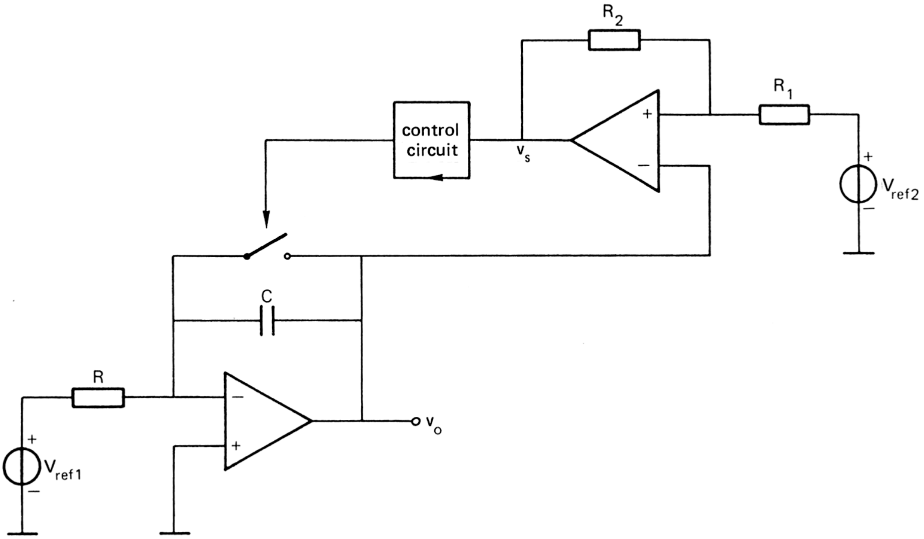


Figure 16.11. A functional diagram for a saw-tooth generator. The output voltage is produced by charging a capacitor with a constant current and discharging it through a switch.

Figure 16.12 shows the various output voltages in the case of a negative reference voltage. The switch goes on as soon as v_o reaches the Schmitt-trigger's upper switching level. The capacitor discharges very rapidly, the output voltage then drops to zero and remains at zero as long as the switch is on. This means that the lower switching level of the Schmitt-trigger (which is βV^- for $V_{ref2} = 0$, see Section 14.2.2) must be higher than zero otherwise the Schmitt-trigger will remain in the $v_s = V^-$ state, the switch will never go off again and the output will stay at zero. This explains the need for the second reference voltage V_{ref2}. The switching levels of the Schmitt-trigger are determined by R_1, R_2 and V_{ref2}.

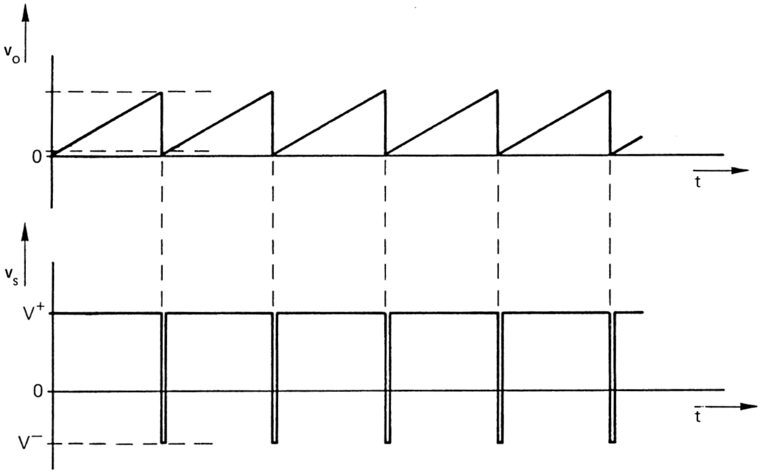


Figure 16.12. The output voltage v_o of the integrator and the output voltage v_s of the Schmitt-trigger in the circuit seen in Figure 16.11. The dotted lines in v_o represent the switching levels of the Schmitt-trigger. For $v_s = V^+$ the switch is on, for $v_s = V^-$ it is off.

Example 16.1

The power supply voltages of the operational amplifiers in Figure 16.11 are $V^+ = 15$ V and $V^- = -15$ V. The output levels of the Schmitt-trigger are assumed to be equal to the supply voltages or, in other words, +15 and –15 V. The switching levels are calculated as follows. Both levels are described as:

$$V_{ref2}\frac{R_2}{R_1+R_2}+v_s\frac{R_1}{R_1+R_2}$$

The upper level is for $v_s = +15$ V, so it amounts to:

$$V_{ref2}\frac{R_2}{R_1+R_2}+15\frac{R_1}{R_1+R_2}$$

which is also the upper ramp peak v_o. The lower switching level is for $v_s = -15$ V, which means that it is:

$$V_{ref2}\frac{R_2}{R_1+R_2}-15\frac{R_1}{R_1+R_2}$$

The minimum value of v_o is zero (the discharged capacitor). To allow the ramp to run from +1 V to +10 V, the upper level must satisfy:

$$V_{ref2}\frac{R_2}{R_1+R_2}+15\frac{R_1}{R_1+R_2}=10$$

and the lower level must be:

$$V_{ref2}\frac{R_2}{R_1+R_2}-15\frac{R_1}{R_1+R_2}=1$$

These equations establish the ratio between R_1 and R_2 as well as V_{ref2}: 3/7 and +7.9 V, respectively.

The width of the pulse-shaped voltage v_s is determined by the discharge time of the capacitor and the delay times of the switch and the Schmitt-trigger. This pulse is, in any case, very narrow thus creating the extremely steep rising edge of v_o. The Schmitt-trigger could be replaced by a comparator (a Schmitt-trigger without hysteresis). However, due to the on-resistance of the switch (Section 15.1), the discharge period is not zero (it is actually an exponentially decaying curve). Due to the Schmitt-trigger hysteresis, the switch will only go off again if v_o is sufficiently close to zero, irrespective of the discharge time.

The frequency of the generated ramp voltage is determined by the time constant RC, by V_{ref1} and by the hysteresis of the Schmitt-trigger. The control circuit adjusts the output of the Schmitt-trigger to the levels appropriate for activating the switches.

16.2.3 Square wave and pulse generators

Most circuits used for the generation of square wave and pulse-shaped signals are composed of resistors, capacitors and a number of digital circuits, they can also be created with an operational amplifier. The principle of operation is essentially the same. Figure 16.13a shows a square wave generator with a Schmitt-trigger. The Schmitt-trigger output (the operational amplifier together with R_1 and R_2) is connected to the input via an integrating RC-network. Again, the capacitor is periodically charged and discharged but this time that is not done with a constant current but via the resistor R instead. The voltage v_c across the capacitor therefore approaches the respective levels V^+ and V^- exponentially. The switching levels are βV^+ and βV^-. The frequency of the generated signal can be derived from Figure 16.13b. It appears that the period time is

$$T=\tau\ln\left(\frac{V^- -\beta V^+}{V^- -\beta V^-}\cdot\frac{V^+ -\beta V^-}{V^+ -\beta V^+}\right) \tag{16.8}$$

with $\tau = RC$.

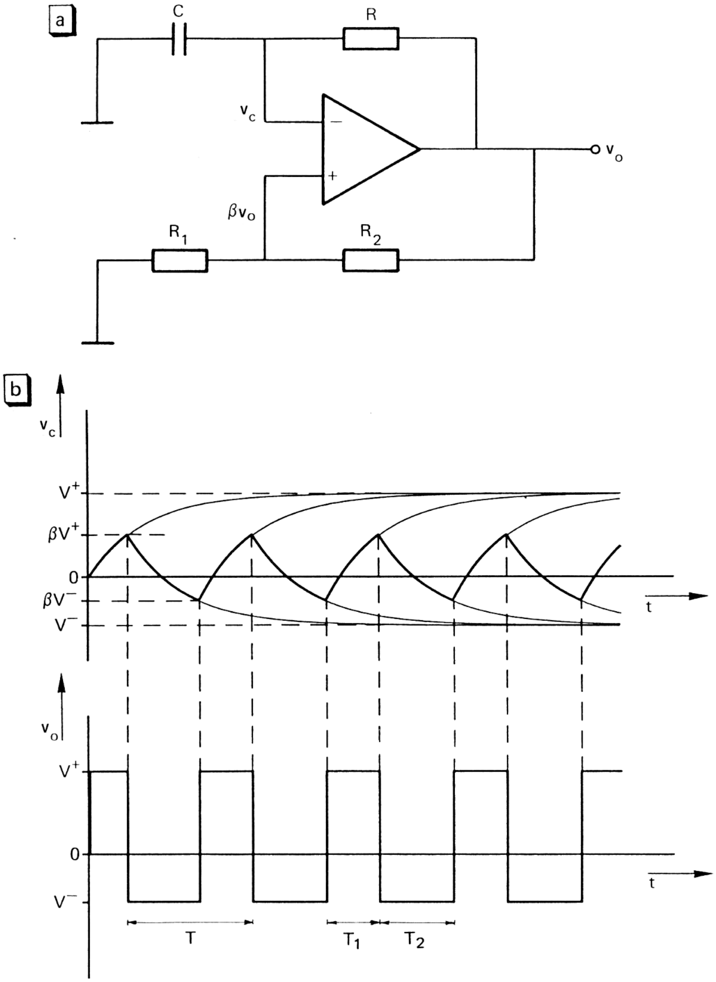


Figure 16.13. (a) The circuit and (b) the corresponding voltages of a square-wave generator constructed with a Schmitt-trigger.

16.2.4 Voltage-controlled oscillators

Many applications require a generator with a frequency that can be controlled by a voltage. Such generators are called voltage-controlled oscillators (VCO) or sweep generators. The latter name is an allusion to the chance to linearly or logarithmically change the frequency of the VCO. The control voltage is a triangular or ramp voltage that sweeps the frequency of the VCO between two adjustable values.

The oscillators of Section 16.1 are not suitable for a VCO because their frequency is determined by resistances and capacitances which can barely be electronically controlled, at least not over a very wide range. The generators with periodically charging and discharging capacitors are better suited to VCOs. The charging time and therefore also the frequency is mainly determined by the charge current. For instance, the frequency of the ramp generator seen in Figure 16.11 is directly proportional to the reference voltage V_{ref1}.

The principle of the VCO is the same as that of the triangle generator seen in Figure 16.8. The frequency of the triangular voltage is proportional to the charge current I_1 and the discharge current I_2. A VCO may be obtained by employing a voltage-to-current converter to produce these currents.

VCOs are available as integrated circuits. Figure 16.14 shows a block diagram with the connections for such circuits. The circuit in question contains two buffered outputs: one for a triangular voltage, the other for a square wave signal. The frequency is controlled by the input voltage v_i. Depending on the type, the sweep range will be a factor of 3 to 10. The sweep range can be changed by external resistances or capacitances within a 1 Hz to 1 MHz range.

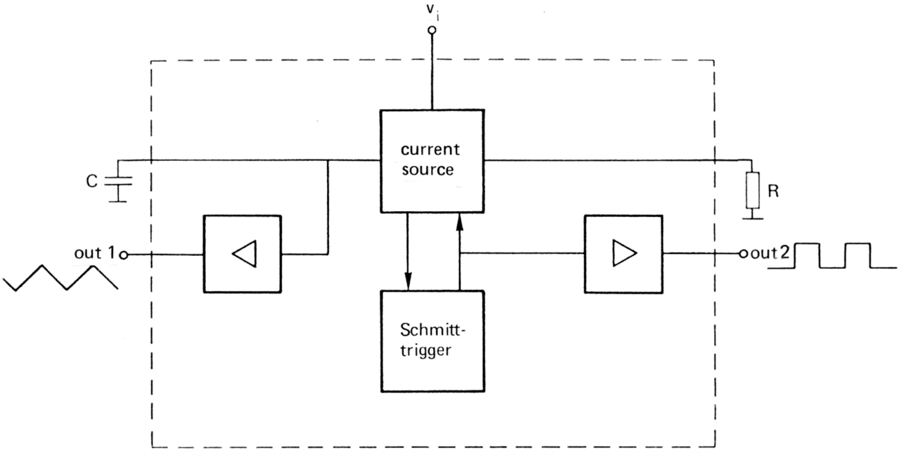


Figure 16.14. An integrated voltage-controlled oscillator (VCO) with triangular and square wave outputs.

There are also VCO types with a much wider sweep range, their output is usually just a pulse or a square wave. The frequency sweep of such kinds of VCOs may be more than a factor of 1000 while the frequency range rises to over 10 MHz.

Such circuits are usually called voltage-to-frequency converters, in particular when there is an accurate relationship between the control (or input) voltage and the output frequency. They are used to convert sensor signals into frequency-analog signals in order to improve noise immunity when transmitting sensor signals over large distances.

SUMMARY

Sine wave oscillators

- One way of generating a sine wave is by using the electronic solution to a second order differential equation which is a sine wave. Other methods involve: filtering the harmonics from a triangular or rectangular signal wave, sine-shaping by means of non-linear transfer circuits and synthesizing with a computer.

- A harmonic oscillator consists of an amplifier, a frequency-selective network and an amplitude control circuit. The parameter that has to be controlled is the damping factor of the second order system.
- The oscillation condition is $A\beta(\omega) = 1$, in which A and β are the amplifier transfers and the feedback network. The solving of this equation creates the conditions for the gain and oscillation frequency.
- The most frequently used types of oscillators are: the Wien oscillator (an amplifier with a Wien network), the phase-shift oscillator (an amplifier with a cascaded series of RC networks) and the two-integrator oscillator or dual integrator loop (comprising a loop of two integrators and an inverting amplifier).

Voltage generators

- A triangular or ramp voltage can be produced by alternatingly charging and discharging a capacitor with a constant current. The switching moments are determined by the Schmitt-trigger switching levels and the peak-to-peak value by its hysteresis interval.
- The duty cycle of a pulse-shaped signal is the ratio between the "on" time and the period time.
- A voltage-controlled oscillator (VCO) is a generator with a frequency that can be varied according to the voltage. The sensitivity of a VCO is given in Hz/V or kHz/V.

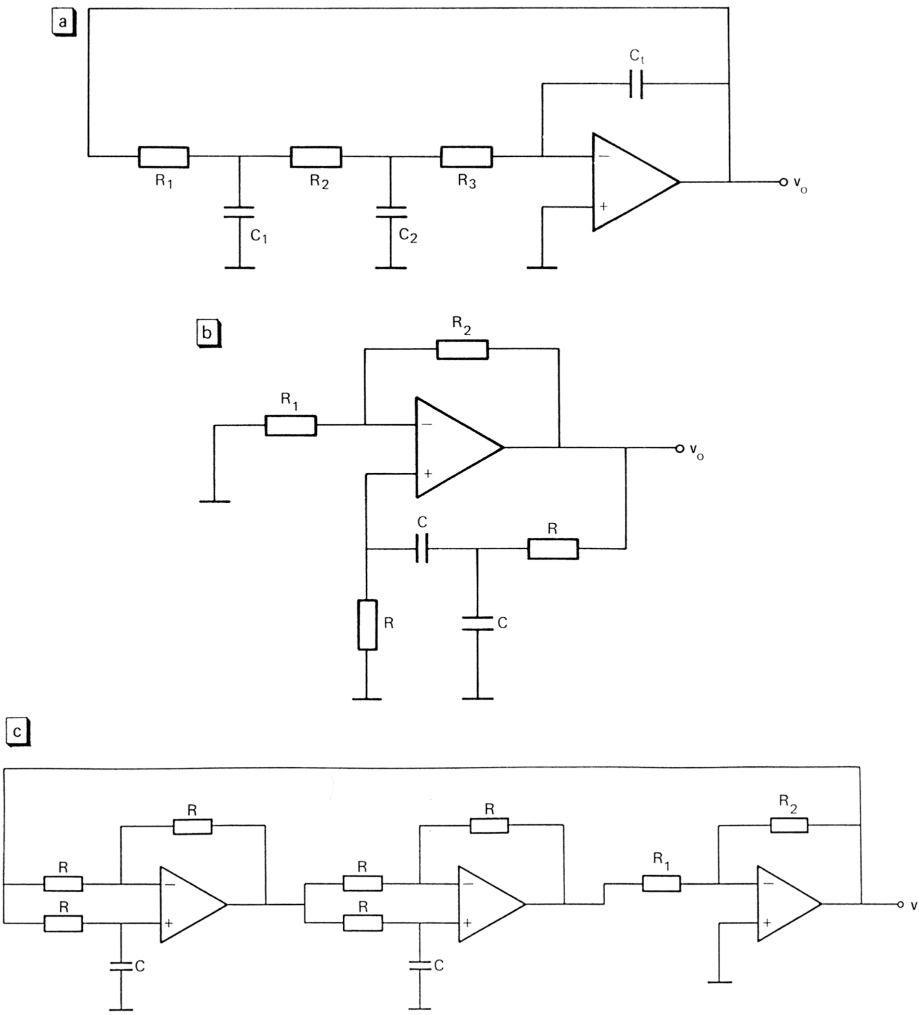
EXERCISES

Sine wave oscillators

16.1 Why does a harmonic oscillator need an amplitude control circuit?

16.2 In $\alpha < 0$ the solution to Equation (16.2) is a sine wave with exponentially increasing amplitude. However, when the system is switched on, the initial amplitude will be zero. Explain why the circuit will start to oscillate?

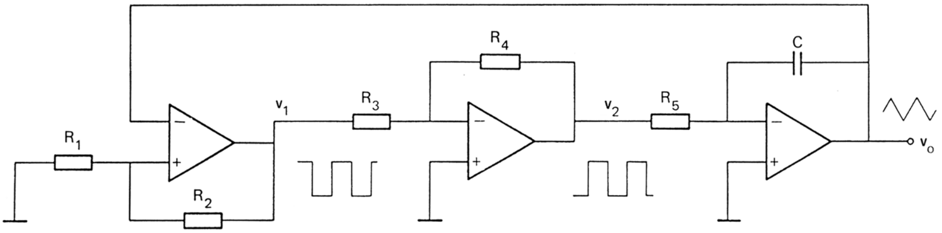
16.3 Find the oscillation conditions for circuits a-c given below. Determine the oscillation frequency and the conditions for the component values. The operational amplifiers may be considered to be ideal. To simplify the calculations, take $R_1 = R_2 = R$; and $C_1 = C_2 = C$.



16.4 Look at the sine-wave oscillator given in Figure 16.4b. The amplitude control is realized by replacing R_4 with a thermistor. The following applied component properties are given: $R_3 = 500\ \Omega$; $R_4 = R_{th} = 10^4 e^{5200(1/T-1/T_0)}$. The heat resistance between the thermistor and the ambient is 100 K/W. $T_0 = 273$ K, the ambient temperature is 300 K. Find the amplitude of the output signal.

Voltage generators

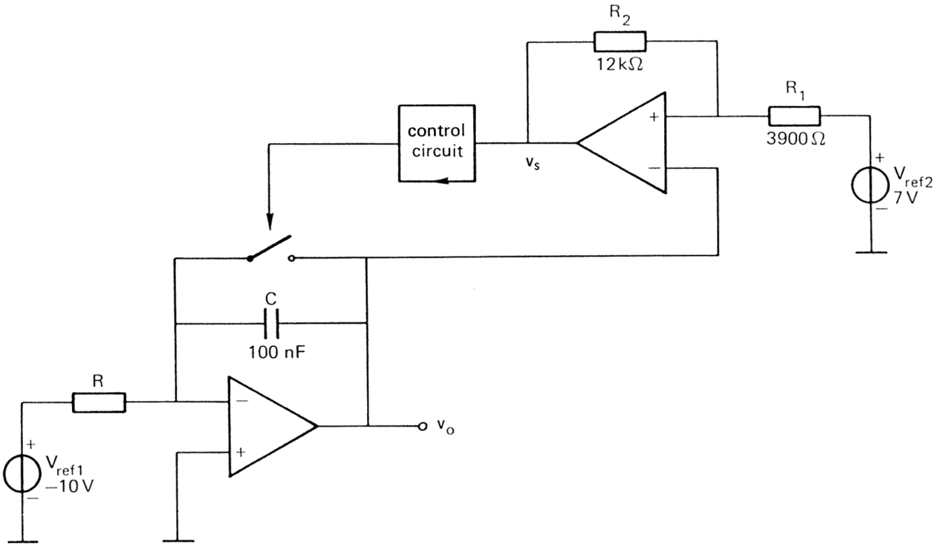
16.5 Imagine that in the case of the triangle generator given below the following conditions hold: $R_3 = 5R_4$, $R_1/(R_1 + R_2) = \beta$ and $R_5C = \tau$. Discuss the changes in v_o when the value of R_1 increases.



16.6 Calculate the duty cycle of the voltages v_1 and v_2 in the Figure accompanying Exercise 16.5, in the case of $V^+ = 15$ V and for $V^- = -5$ V.

16.7 Calculate the frequency, average value and amplitude of the triangular voltage v_o in the circuit given in Exercise 16.5 when $V^+ = 15$ V, $V^- = -5$ V, $R_5 = 10$ kΩ, $C = 100$ nF and $\beta = 0.5$.

16.8 Examine the ramp generator depicted below. The output frequency of v_o is 1 kHz. Establish the value of R. All components may be considered ideal and the output levels of the Schmitt-trigger are $v_s = V^+ = 18$ V or $v_s = V^- = -12$ V.



16.9 Study the generator in the preceding exercise. By adding only one voltage source the average output voltage can be adjusted to zero. How can it be adjusted and what is the value?

16.10 Look again at the ramp generator in Exercise 16.8. Find the amplitude and the frequency of the output signal v_0 for $R = 100$ kΩ, the switch's delay time is 2 ms but apart from that the switch is ideal.

16.11 Modify the circuit in Exercise 16.8 so that the output has a negative slope.

16.12 Determine the duty cycle of the output signal in Figure 16.13 as a function of V^+, V^-, β and $\tau = RC$. Find the condition for a 50% duty cycle.

17 Modulation and demodulation

Modulation is a special kind of signal conversion that makes use of an auxiliary signal known as the carrier. One of the parameters of this particular carrier is varied in proportion to the input or measurement signal, all of which results in a shift in the complete signal frequency band over a distance related to the carrier frequency. It is because of this property that modulation is sometimes also termed frequency conversion.

The modulated waveform has a number of advantages over the original waveform and that is why it is widely used in various electrical systems. One modulation application is frequency multiplexing which is one way of transmitting signals more efficiently (see Chapter 1.1). Telecommunication would be inconceivable without modulation. One major advantage of modulated signals is that they provide better noise and interference immunity. It is particularly in instrumentation systems that modulation makes it possible to bypass offset and drift. Indeed, that is the main reason why we want to discuss modulation in this book.

The carrier is a simple waveform signal like, for instance: a sine wave, a square wave, or a pulse-shaped signal. Several carrier parameters can be modulated using the input signal, examples being: the amplitude, the phase, the frequency, the pulse height or the pulse width. These types of modulation are known as amplitude modulation (AM), phase modulation, frequency modulation (FM), pulse height and pulse width modulation. Figure 17.1 shows some examples of this, in the figure x_i is the modulator input signal, x_d the carrier signal and x_o the output or modulated signal.

With FM signals (Figure 17.1d) the original information is embedded in the zero crossings of the modulated signal whereas in an amplitude modulated signal the information tends to be included in the peak values. These kinds of peak values are easily affected by additive interference, the zero position crossings are much less sensitive to interference which is why FM signals have better noise immunity than AM signals. Nevertheless, even AM is a powerful tool in instrumentation for suppressing interference. This chapter will therefore focus exclusively on amplitude modulation. The first part will deal with the basic concepts of amplitude modulation,

modulation methods and the reverse operation which is demodulation. In the second part of the chapter some of the applications of modulation and demodulation in particular measurement instruments will be described.

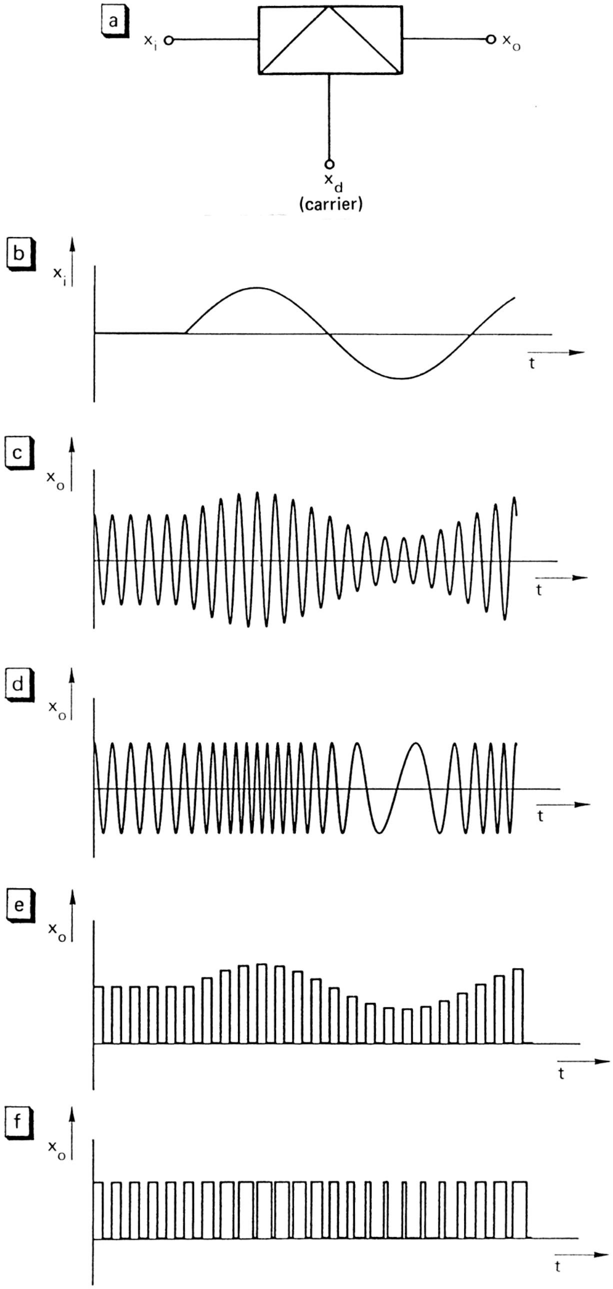


Figure 17.1. (a) The circuit symbol for a modulator, (b) an example of an input signal, (c) amplitude modulation, (d) frequency modulation, (e) pulse height modulation, (f) pulse width modulation.

17.1 Amplitude modulation and demodulation

17.1.1 Theoretical background

The properties of modulated signals will be analyzed by describing time domain signals. We shall start with a sinusoidal carrier $x_d(t) = \hat{x}_d \cos\omega_d t$. The amplitude of the carrier will vary according to the input signal $x_i(t)$ and will result in the time-varying carrier amplitude $\hat{x}_d(t) = \hat{x}_d(1+kx_i(t))$. The modulated output signal may therefore be given as:

$$x_o(t) = \hat{x}_d(t)\cos\omega_d t = \hat{x}_d\bigl(1 + kx_i(t)\bigr)\cos\omega_d t \tag{17.1}$$

in which k is a coefficient determined by the modulator. With $x_i(t) = 0$, the output is merely $x_d(t)$, which is the original carrier signal with constant amplitude. Let us suppose that the input signal is a pure sine wave with only one frequency ω_i: $x_i(t) = \hat{x}_i \cos\omega_i t$. If that is so then the modulated signal will be:

$$x_o(t) = \hat{x}_d\bigl(1 + kx_i \cos\omega_i t\bigr)\cos\omega_d t \tag{17.2}$$

If this expression is expanded into sine-wave components it will result in:

$$\begin{aligned} x_o(t) &= \hat{x}_d\bigl(\cos\omega_d t + m\cos\omega_i t \cos\omega_d t\bigr) = \\ &= \hat{x}_d\Bigl(\cos\omega_d t + \tfrac{1}{2}m\cos(\omega_d + \omega_i)t + \tfrac{1}{2}m\cos(\omega_d - \omega_i)t\Bigr) \end{aligned} \tag{17.3}$$

where $m = k\hat{x}_i$ is the modulation depth, a term that is apparent from Figure 17.2. This figure represents the output waveform for two different values of m. The input signal can still be recognized in the envelope of the modulated signal.

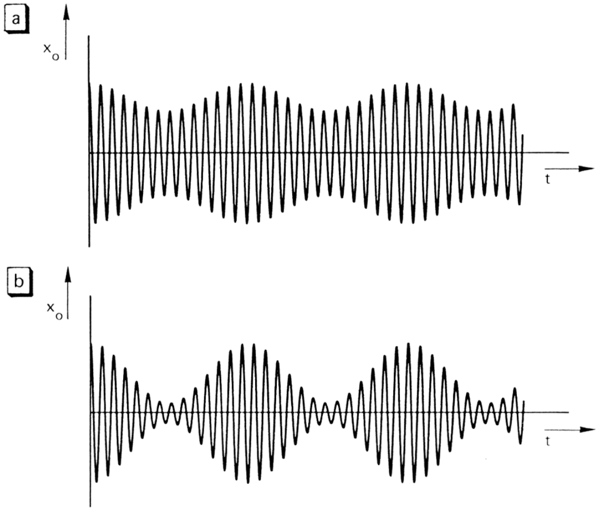


Figure 17.2. A sinusoidal carrier modulated by a sinusoidal input signal with modulation depths of (a) m = 0.25 and (b) m = 0.75.

Evidently the modulated signal has three frequency components: one with the carrier frequency (ω_d), one with a frequency that is equal to the sum of the carrier frequency and the input frequency ($\omega_d + \omega_i$), and one that encompasses the difference between these two frequencies ($\omega_d - \omega_i$) (Figure 17.3a).

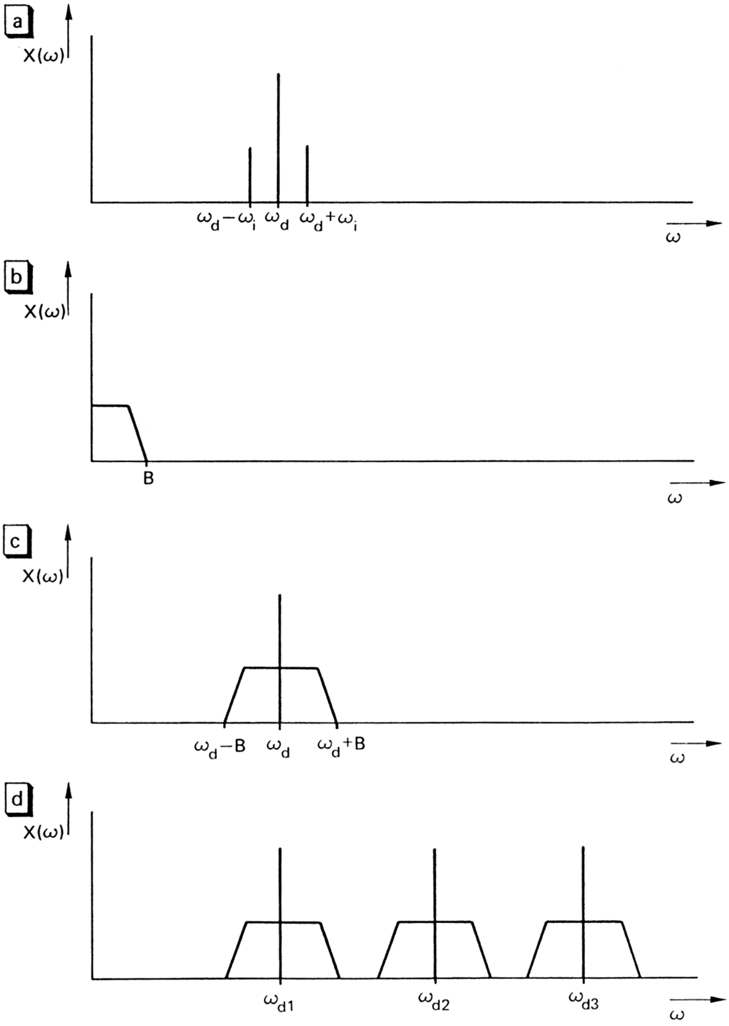


Figure 17.3. The frequency spectra of (a) a sine wave carrier (ω_d) modulated with a sine wave (ω_i), (b) a low frequency input signal with bandwidth B, (c) the corresponding AM signal, (d) a frequency-multiplexed signal.

The modulator produces two new frequency components for each input component that are positioned on either side of the carrier frequency. An arbitrary, aperiodic low-frequency input signal has a continuous spectrum as depicted in Figure 17.3b. When the amplitude of a carrier is modulated using this signal the whole frequency band will be shifted to a region around the carrier frequency (Figure 17.3c). Both bands on either side of the carrier are called the modulated signal's side bands. The total bandwidth of an AM signal thus becomes twice that of the original signal. Each side band carries all the information contained in the input signal.

From the foregoing discussion it is clear that – if the carrier frequency is high enough – an AM signal will not contain a DC component or low frequency components, even when the original input signal does indeed comprise such components. Modulated signals can therefore be amplified without being disturbed by offset and drift: such error signals can easily be removed from the amplified output with the help of a high-pass filter.

The new position of the frequency band lies just around the carrier frequency. The frequency band of another measurement signal can be moved to a different position by using an alternative frequency to modulate the carrier (Figure 17.3d). If these bands, and possibly other bands as well, do not overlap they can all be transported over a single transmission channel (a cable, a satellite link) without disturbing each other. Once they have been received, the signals are separated by demodulation and returned to their original position in the frequency band.

Example 17.1

The acoustic signal bandwidth in a telephone system is limited to a range of 300 to 3400 Hz. By contrast, the transmission channels have a much wider bandwidth. To make signal transport effective the various acoustic signals have to be converted to different carrier frequencies so that many signals can be simultaneously transported over one conductor pair. As both side bands contain the same information a special modulation technique known as single sideband modulation is introduced thus making it possible to double the channel capacity. To further improve transmitter efficiency the carrier frequency is not transported by the signals.

The first signal is converted to a band between 12 and 16 kHz, the second to a band between 16 and 20 kHz, and so on. In this way one cable consisting of 24 conductor pairs can carry 2880 telephone signals.

17.1.2 Amplitude modulation methods.

There are many ways to modulate the amplitude of a carrier signal. One way is by employing an analog multiplier (Chapter 14). When a carrier signal $x_d(t) = \hat{x}_d \cos\omega_d t$ is multiplied by an input signal $x_i(t)$ this will result in an output signal $x_o(t) = x_i(t) \cdot x_d(t) = x_i(t) \cdot \hat{x}_d \cos\omega_d t$. For simplicity's sake, the scale factor of the multiplier is set at 1. With a sine-shaped input signal $x_i(t) = \hat{x}_i \cos\omega_i t$ the output of the multiplier is

$$x_o(t) = \hat{x}_i \hat{x}_d \cos\omega_i t \cos\omega_d t = \tfrac{1}{2}\hat{x}_i \hat{x}_d \left\{\cos(\omega_d + \omega_i)t + \cos(\omega_d - \omega_i)t\right\} \qquad (17.4)$$

This signal just contains the two side band components but not the carrier (see Figure 17.4a). Figure 17.4b shows this signal in the time domain. Because there is no carrier it is called an AM signal with a suppressed carrier. In the case of arbitrary input signals the spectrum of the AM signal consists of two (identical) side bands without a carrier.

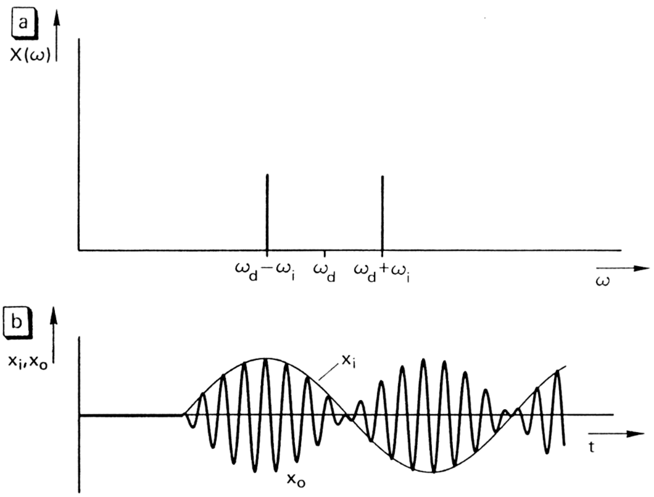


Figure 17.4. (a) The frequency spectrum and (b) an AM signal amplitude-time diagram with a suppressed carrier.

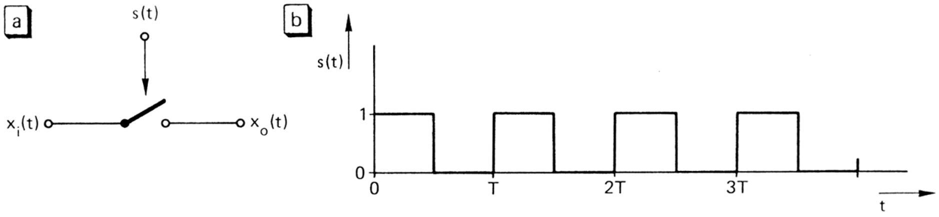


Figure 17.5. (a) An electronic switch that can act as a modulator, (b) the switching signal *s*(*t*) that serves as a carrier.

A second type of modulator is the switch modulator in which the input signal is periodically switched on and off, a process that can be described by multiplying the input signal by a switch signal $s(t)$ that is 1 when the switch is on and 0 when the switch is off (Figure 17.5). This results in a modulated signal $x_o(t) = x_i(t)s(t)$. To show that this product is indeed a modulated signal we can expand $s(t)$ to its Fourier series:

$$s(t) = \tfrac{1}{2} + \tfrac{2}{\pi}\left\{\sin\omega t + \tfrac{1}{3}\sin 3\omega t + \tfrac{1}{5}\sin 5\omega t + \ldots\right\} \tag{17.5}$$

(see example 17.5)
with $\omega = 2\pi/T$, and T as the switching signal period. With a sine wave input signal $x_i(t) = \hat{x}_i \cos\omega_i t$, the output signal $x_o(t)$ is:

$$x_o(t) = \hat{x}_i\left\{\tfrac{1}{2} + \tfrac{2}{\pi}\sum_{n=1}^{\infty}\tfrac{1}{n}\sin n\omega t\right\}\cos\omega_i t =$$

$$= \tfrac{1}{2}\hat{x}_i\cos\omega_i t + \tfrac{2}{\pi}\hat{x}_i\left[\tfrac{1}{2}\left\{\sin(\omega+\omega_i)t + \sin(\omega-\omega_i)t\right\} + \tfrac{1}{2}\cdot\tfrac{1}{3}\left\{\sin(3\omega+\omega_i)t + \sin(3\omega-\omega_i)t\right\}\ldots\right] \tag{17.6}$$

with n even. The spectrum of this signal is depicted in Figure 17.6a while Figure 17.6b shows the spectrum for an arbitrary input signal.

This modulation method produces a large number of side band pairs positioned around the fundamental and odd multiples (ω, 3ω, 5ω, ...). The low-frequency component derives from multiplying the mean of $s(t)$ (here it is $\frac{1}{2}$). This low-frequency component and all components with frequencies 3ω and higher can be removed with the help of a band-pass filter. The resulting signal will be just an AM signal with a suppressed carrier (Figure 17.6c).

The advantages of the switch modulator are its simplicity and its accuracy. The side band amplitude is only determined by the quality of the switch. A similar modulator can be created by periodically changing the polarity of the input signal. This will be equivalent to multiplying it by a switch signal that has zero as the mean value. In such cases there is no low-frequency component as in Figure 17.6.

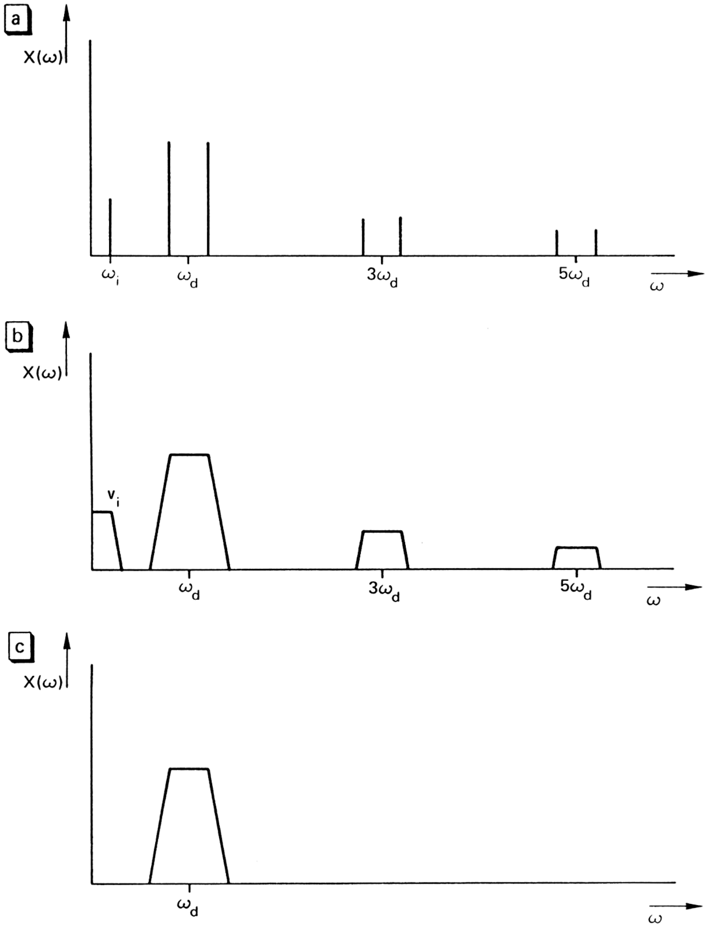


Figure 17.6. Signal spectra for a switch modulator, (a) the output spectrum for a sinusoidal input signal, (b) the output spectrum for an arbitrary input signal with a limited bandwidth, (c) as in (b) after low-pass filtering.

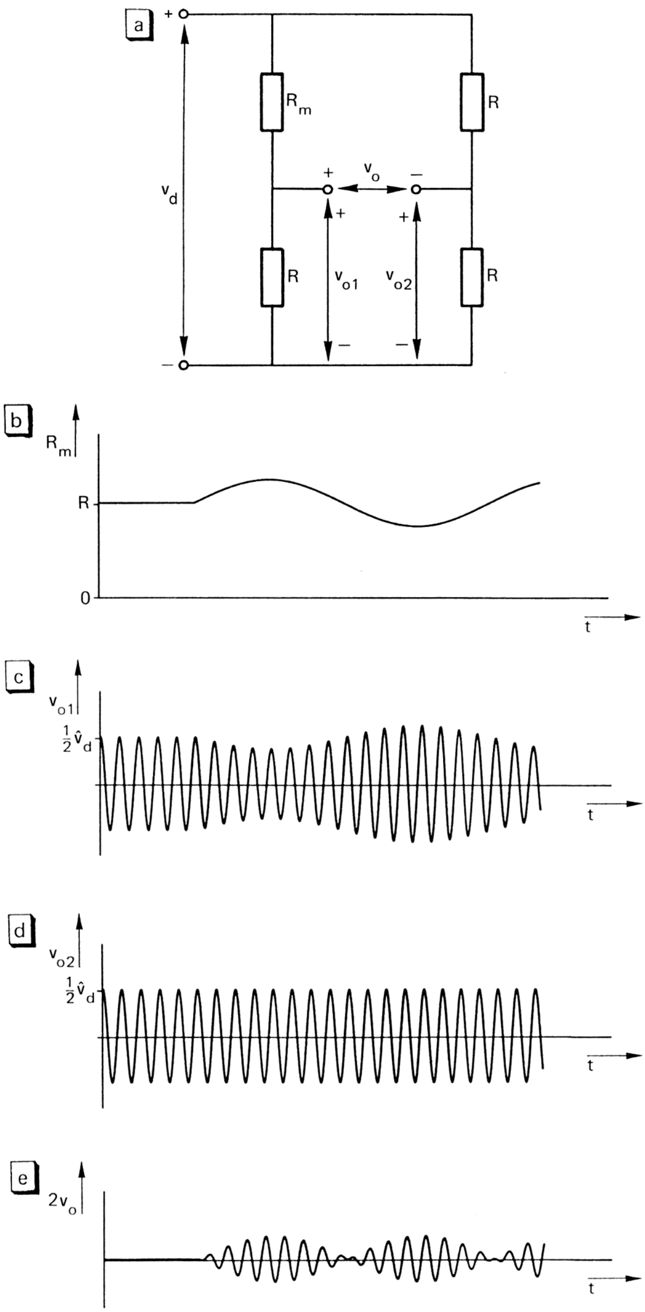

a
+
−
v_d
R_m
R
R
R
+
v_o
−
+
+
v_o1
v_o2
−
−
b
R_m
R
0
t
c
v_o1
½ v̂_d
t
d
v_o2
½ v̂_d
t
e
2v_o
t

Figure 17.7. (a) A resistance measurement bridge for the measurement of resistance R_m,(b) an example of varying resistance value, (c) v_{o1} is an AM signal with the carrier $\frac{1}{2}v_d$,(d) v_{o2} has a constant value of $\frac{1}{2}v_d$,(e)the bridge output voltage $v_{o1} - v_{o2}$: an AM signal without a carrier (scaled to a factor of 2 in the figure).

We have seen that the absence of DC and low-frequency components considerably encourages the amplification of modulated signals so that offset, drift and low frequency noise can be kept far from the signal frequency band. When very low voltages need to be measured, and thus amplified, it is wise to modulate them before continuing with other types of analog signal processing that might introduce DC errors. One application of this concept is encountered in the measurement bridge which may be viewed as a third modulation method. The principle is well illustrated by the resistance measurement bridge or the Wheatstone bridge given in Figure 17.7a. Such bridge circuits are widely used in the readout of resistive sensors where resistance values change according to the physical quantity applied (Section 7.2). The bridge is connected to an AC signal source and the AC signal (usually a sine or square wave) acts as the carrier. In the example given here there is just one varying resistance R_m in the bridge (Figure 17.7b). If the bridge is conceived as a double voltage divider then the output voltage of the left-hand branch will be found to be $v_{o1} = v_d R/(R + R_m)$ while the right-hand branch output is $\text{v}_{o2} = \frac{1}{2} v_d$. The latter voltage has a constant amplitude (see Figure 17.7d) while the amplitude of v_{o1} varies with R_m. It is an AM signal modulated by R_m (Figure 17.7c) and it may also be written as $v_{o1} = \frac{1}{2} v_d(1 + f_r(t))$, in which $f_r(t)$ varies in proportion to R_m. The bridge output signal is $v_o = v_{o1} - v_{o2} = \frac{1}{2} v_d(1 + f_r(t)) - \frac{1}{2} v_d = \frac{1}{2} f_r(t) v_d$, which is an AM signal without a carrier (suppressed by subtracting v_{o2}), as shown in Figure 17.7e. This output can be amplified by introducing a differential amplifier with a high gain, its low frequency properties are irrelevant. The only requirements are a sufficiently high bandwidth and a high CMRR for the carrier frequency so that the $v_{o1} - v_{o2}$ difference can be accurately amplified.

17.1.3 Demodulation methods

The opposite process to modulation, sometimes misnominously termed detection, is demodulation. If we consider the AM signal and its carrier (as seen, for instance, in Figure 17.2a) we may observe the similarity between the amplitude envelope and the original signal shape. One obvious demodulation method is peak detection (see Section 9.2.2). Although this method is common in AM signal radio receivers, it is not recommended for instrumentation purposes because of its inherent inaccuracy and noise sensitivity: each transient is taken as a new peak belonging to the signal. A better demodulation method would be double-sided rectifying and low-pass filtering (see Figure 17.8).

The low-pass filter responds to the rectified signal average. When properly designed, the output follows the input amplitude.

Obviously the peak detector and the rectifier detector only operate with AM signals that have carriers. In the case of AM signals without carriers the envelope is no longer a copy of the input, there the positive and negative input signals produce equal amplitudes (Figure 17.4b). Clearly full recovery of the original waveform requires additional information with respect to the input phase.

An excellent way of solving this problem, and it is a way that has a number of additional advantages, is by introducing synchronous detection. This method involves multiplying the AM signal by another signal that has the same frequency as the original carrier. If the carrier signal is available (as is the case in most measurement systems) this signal can be the carrier itself.

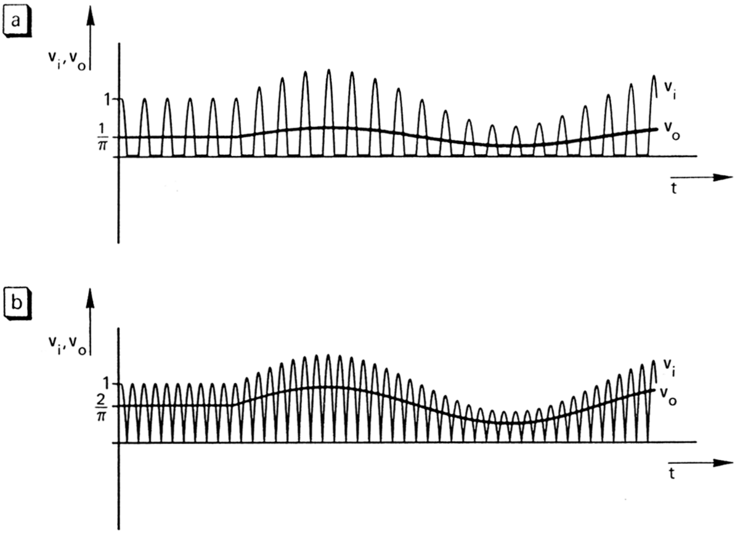


Figure 17.8. (a) A single-sided rectified AM signal v_i followed by a low-pass filter that responds to the average value v_o, (b) like (a) but with a double-sided rectified signal to produce an average that is twice as high.

Let us suppose that we have an AM signal with a suppressed carrier: $x_o = \hat{x}_o \cos\omega_d t \cos\omega_i t$. This is then multiplied by the synchronous signal $x_s = \hat{x}_s \cos(\omega_d t + \varphi)$, where φ takes into account a possible phase shift relative to the original carrier. The product of these signals is $x_{dem} = \hat{x}_s \hat{x}_o \cdot \cos\omega_d t \cdot \cos\omega_i t \cdot \cos(\omega_d t + \varphi)$. If the frequency components are separated this will result in:

$$x_{dem}(t) = \tfrac{1}{2}\hat{x}_o\hat{x}_s \cos\omega_i t \cos\varphi + \tfrac{1}{4}\hat{x}_o\hat{x}_s\left[\cos\{(2\omega_d + \omega_i)t + \varphi\} + \cos\{(2\omega_d - \omega_i)t + \varphi\}\right] \tag{17.7}$$

With a low-pass filter the components around $2\omega_d$ will be removed, thus leaving us with the original component that has a frequency of ω_i. In case x_s is in phase with the carrier ($\varphi = 0$), this component has a maximum value. The component is zero when the phase difference amounts $\varphi = \pi/2$; the component gets the negative value for a phase difference equal to $\varphi = \pi$. This phase sensitivity is an essential property of synchronous detection. Figure 17.9 reviews the whole detection process.

Figure 17.9b depicts the spectrum of an AM signal without a carrier, together with certain error signal components. If it is multiplied by the synchronous signal a new band will be created that coincides with the original band (Figure 17.9c). The low-pass filter removes all components with frequencies that are higher than those of the original band (Figure 17.9d).

One important advantage of this detection method is that it helps to eliminate all error components not present in the AM signal's small band. If the measurement signal has a narrow band (with slowly fluctuating measurement quantities), then a low filter frequency cut-off can be chosen. Most of the error signals can therefore be removed, even with a simple first order low-pass filter. Synchronous detection facilitates AC-signal measurement with very low signal-to-noise ratios.

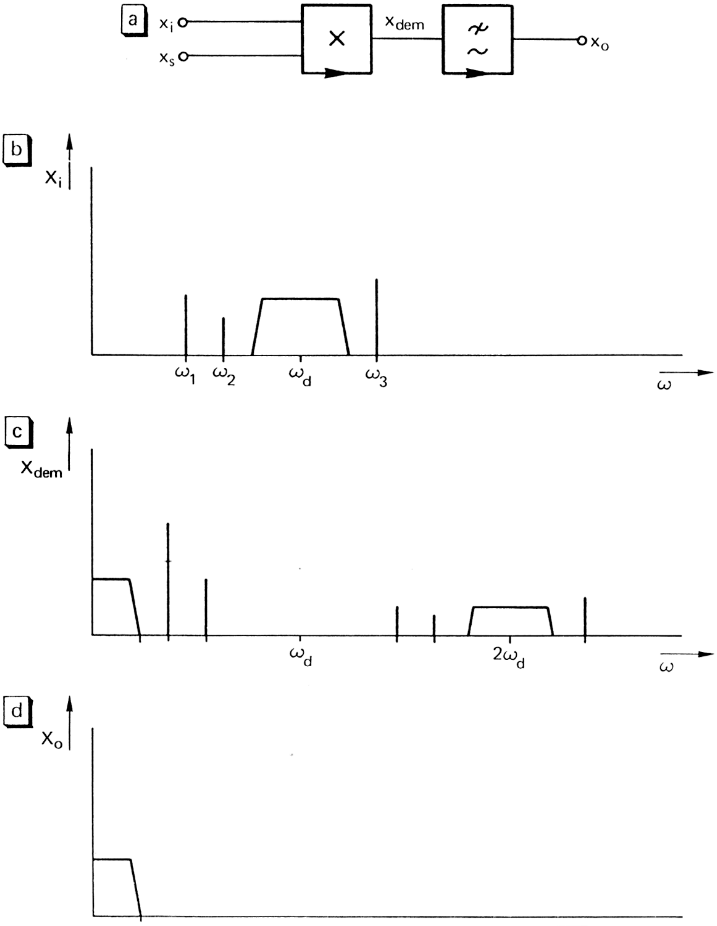


Figure 17.9. (a) A synchronous detector composed of a multiplier and a low-pass filter, (b) the frequency spectrum of an AM signal without a carrier but with some error signals, (c) the spectrum after synchronous signal multiplication, (d) output after low-pass filtering: the original signal component.

17.2 Systems based on synchronous detection

The preceding section showed that synchronous detection is a powerful instrumentation mechanism for the measuring of small AC signals with low signal-to-noise ratio. The measurement bridge discussed in Section 17.1.2 is supplied by an AC source and, in order to bypass the offset and drift problems of high-gain amplifiers, the low frequency sensor signal is converted into an AM signal. Synchronous detection of amplified bridge output makes the measurement extremely tolerant to noise and other interference components (provided that they remain outside the signal band). The bandwidth of the detection system is set by a simple first order low-pass filter.

Many measurement instruments make use of synchronous detection, examples being: network analyzers and impedance analyzers. In such instruments the measurement signals are all sinusoidal, in other words, the analysis takes place in the frequency domain. Where there is no synchronous signal it first has to be generated. This is done by making use of a special system called the phase-locked-loop that will be described in the next section. Synchronous detection is also used in special types of amplifiers, like with the lock-in amplifier and the chopper amplifier. Their principles will be explained in later sections.

17.2.1 The phase-locked loop

A phase-locked loop or a PLL is a special electronic system that is capable of tracing the frequency of a sinusoidal measurement signal despite noise and other spurious signals. The PLL consists of a voltage-controlled oscillator (i.e. a VCO, see Section 16.2.4), a synchronous detector and a control amplifier (Figure 17.10). The output frequency of the VCO is controlled so that it remains equal to that of a particular input signal frequency component. Because of the high operational amplifier gain, the feedback loop forces the amplifier input down to zero. The synchronous detector responds to the frequency difference between the input signal v_i and the VCO output signal v_o as these signals are multiplied. When the frequencies are equal the synchronous detector responds to the phase difference. If the difference is not zero then the control amplifier will control the VCO via v_s in such a way that it will be reduced. In the steady state the amplifier input will be zero and so the synchronous detector output will also be zero. This situation corresponds to equal frequencies and to the phase difference $\pi/2$ between v_i and v_o.

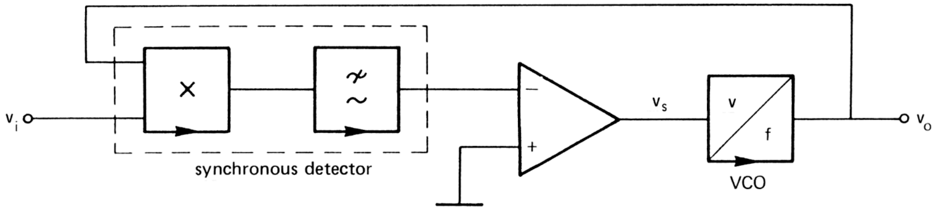


Figure 17.10. A functional diagram of a phase-locked loop (PLL) consisting of a synchronous detector, an amplifier and a voltage-controlled oscillator (VCO).

The synchronous detector is only sensitive to signal frequencies between $f_o \pm B$ (f_o is the fundamental frequency of v_o, and B is the low-pass filter bandwidth). Frequencies outside this band are ignored. The PLL is therefore able to generate a periodic signal with a frequency that is exactly equal to that of one particular input signal component. This property also makes the PLL suitable as an FM-demodulator: the VCO frequency is related to the control voltage v_s. At varying input frequencies v_s varies proportionally: v_s is the demodulated FM signal.

If the circuit given in Figure 17.10 is slightly extended then it becomes possible to generate a periodical signal with a frequency that is equal to the sum of two frequencies (Figure 17.11). Let us now assume that the respective frequencies of the input signals v_1 and v_2 are f_1 and f_2. Synchronous detector 1 will measure the difference between f_1 and f_o which is the frequency of the VCO. The bandwidth of this synchronous detector is chosen in such a way that $f_o - f_1$ can pass the detector but not the sum of $f_o + f_1$. The second detector will measure the difference between the frequencies $(f_o - f_1)$ and f_2. The feedback converts this signal to zero so that $f_2 = f_o - f_1$ or $f_o = f_1 + f_2$.

Likewise, the PLL can be used to generate signals with frequencies that are multiples of the frequency of the input signal. This is achieved by inserting a frequency divider between the VCO output and the synchronous detector input as illustrated in Figure 17.10. If n is the dividend, the synchronous detector measures the difference between f_i and f_o/n, this difference is then converted to zero by the feedback so that $f_o = nf_i$.

The important PLL parameters are the lock-in range and the hold range. The lock-in range represents the range of input frequencies for which the controller automatically locks to that specific frequency. The hold range is the range over which the input frequency may vary after having been locked. Usually the hold range is greater than the lock-in range. The hold range depends on the amplitude of the input signal component: the lower the signal, the more difficult it will be for the PLL to control it. The lock-in range of a PLL can be shifted by changing the free running frequency f_0 (at $v_s = 0$) of the VCO.

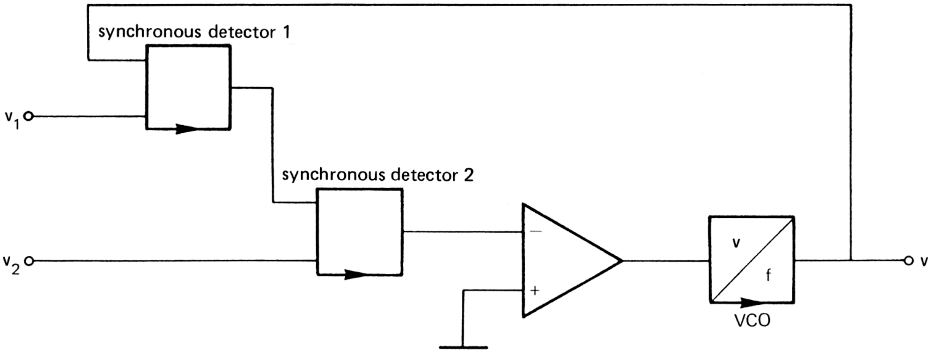


Figure 17.11. An extended PLL to produce a signal with a frequency equal to the sum of the frequencies of v_1 and v_2.

Like many other electronic processing circuits, the PLL is also available as an integrated circuit. This IC contains the analog multiplier, the control amplifier and the VCO. The low-pass filter of the synchronous detector and the components that fix the free-running frequency of the VCO have to be connected externally because such components are difficult to integrate. A degree of design freedom is also reserved for the user. In the next two sections the PLL will be considered as a complete system (the black box approach).

17.2.2 Lock-in amplifiers

A lock-in amplifier is an AC amplifier that is based on synchronous detection and is intended for the measuring of the amplitude and phase of small, noisy, narrow-band measurement signals. A simplified block diagram of a lock-in amplifier is depicted in Figure 17.12. The amplifier has two input channels: the signal channel and the reference channel. The signal channel consists of AC amplifiers and a band-pass filter. This filter, known as the predetection filter, is used to remove some of the input noise prior to detection. The filter used will be manually adjustable or automatic, all depending on the type of lock-in amplifier that is being used. The reference channel will be composed of an amplifier, an adjustable phase shifter and a comparator. An adjustable low-drift DC amplifier will raise the demodulated signal to the proper output level. The output signal is a DC or a slowly varying signal proportional to the amplitude of the input signal. The adjustable phase shifter allows the amplifier's sensitivity to be maximized: the sensitivity is maximal for $\cos\varphi = 0$ or $\varphi = \pi/2$. The original phase difference between v_i and v_{ref} can be determined with the calibrated phase shifter.

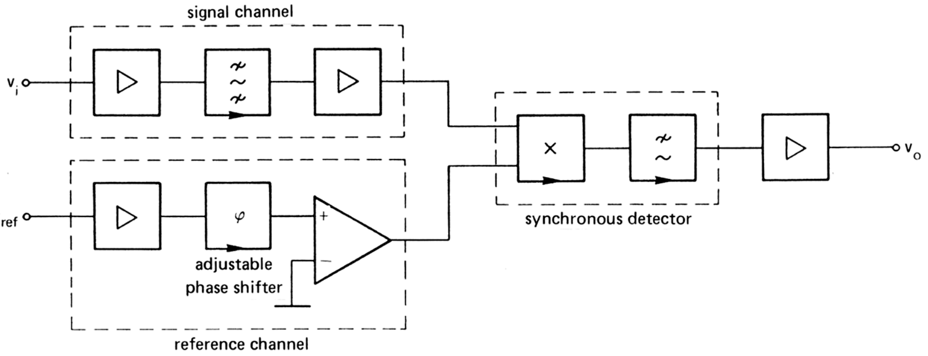


Figure 17.12. A simplified block diagram of a lock-in amplifier used to measure the amplitude and phase of small AC signals.

To properly control the switches in the synchronous switching detector the reference channel comparator converts the reference signal into a square wave. A lock-in amplifier of the type described here is suitable for the measurement of AC signals where synchronous signals are available (just as with bridge measuring circuits or transfer measuring). If there is no such synchronous signal it must be produced. The diagram given in Figure 17.13, where a PLL is used to generate the synchronous signal, explains this process.

17.2.3 Chopper amplifiers

A chopper amplifier, or chopper stabilized amplifier, is a special type of amplifier used for very small DC voltages or low frequency signals. To get rid of the offset and drift encountered in common DC amplifiers, the measurement signal is first modulated using a switch modulator. After that it is amplified and it is finally demodulated by synchronous detection (Figure 17.14).

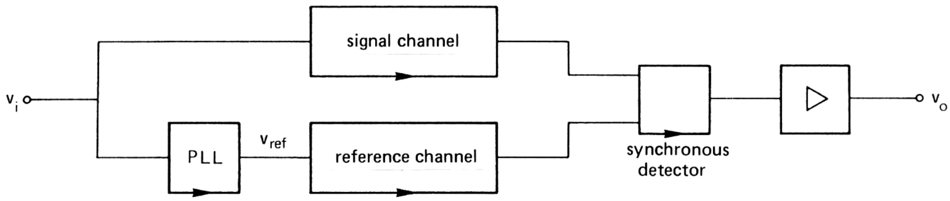


Figure 17.13. A simplified block diagram of a lock-in amplifier with a self-generating reference signal.

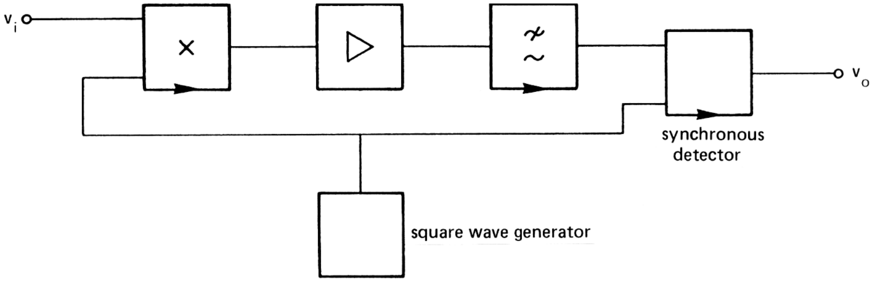


Figure 17.14. A block diagram of a chopper amplifier used for the measurement of very small DC signals.

Chopper amplifiers (also, for obvious reasons, sometimes called indirect DC amplifiers) are available as complete modules. Table 17.1 provides a comparison between the properties of an operational amplifier with JFETs (column I) and two types of chopper amplifiers.

With the column II type the input voltage is chopped using JFETs as switches. With the varactor amplifier shown in column III modulation is performed by varactors, which means to say that reverse biased diodes are used as voltage controllable capacitances. The main advantage of the chopper amplifier is that it combines low input offset voltage with low bias current and the non-zero values are due to imperfections in the switches. The varactor amplifier (sometimes also called an electrometer amplifier after its electrometer tube forerunner) combines very high input impedance with extremely low bias current.

Table 17.1 Selected JFET operational amplifier specifications compared with two types of chopper amplifiers

Parameter	I: JFET	II: JFET chopper	III: varactor
A_0	$2 \cdot 10^5$	10^7	10^5
V_{off}	<0.5 mV	<0.02 mV	-
	adjustable	adjustable	adjustable
t.c. V_{off}	2 - 7 µV/K	0.003 µV/K	10 µV/K
V_{off} supply voltage coefficient	-	±0.1 µV/V	500 µV/V
V_{off} long-term stability	-	±1 µV/month	100 µV/month
I_{bias}	10 – 50 pA	2 pA	0.01 pA
t.c. I_{bias}	2 pA/K	0.5 pA/K	×2 per 10 K
Bandwidth	1 – 4 MHz	2 MHz	2 kHz
R_{i}	10^{12} Ω	10^6 Ω	10^{14} Ω

SUMMARY

Amplitude modulation and demodulation

- Amplitude modulation (AM) is a type of signal conversion in which the amplitude of a carrier signal (which is usually sinusoidal) varies in proportion to the input signal. The carrier frequency is high compared to the input frequency. Modulation leads to changes in the input frequency band.
- The opposite process to modulation is demodulation or detection which results in the original waveform.
- The spectrum of an AM signal consists of two side bands symmetrically positioned around the carrier frequency. The information content in each of these side bands is identical to that of the original band.
- Frequency multiplexing involves multiplexing in such a way that the various signals are converted into different frequency bands by means of modulation. If the bands do not overlap they can be simultaneously transported over a single transmission line.
- Some amplitude modulation methods are:
    - analog multiplying of the input signal and the carrier;
    - periodic switching (on/off or +/–) of the signal, the switching signal acts as the carrier;
    - using a Wheatstone measurement bridge with an AC signal as its carrier.
- Multiplying two signals having frequencies f_1 and f_2 will result in these two new frequencies: the sum of the frequencies $f_1 + f_2$ and the difference between them $f_1 - f_2$.
- Modulated signals are not sensitive to low-frequency interference in conjunction with the position of the frequency band. Small low-frequent measurement signals should, if possible, be modulated prior to any other signal processing. An example of this is the measurement bridge with AC supply voltage.
- Synchronous detection demodulation makes very low signal-to-noise ratio signal measurement possible. A synchronous detector has a phase-sensitive response.
- Diode peak detector demodulation is simple but inaccurate and not applicable to AM signals without carriers.

Systems based on synchronous detection

- Synchronous detection is a type of phase-sensitive amplitude detection suited to the processing of signals with very low signal-to-noise ratios.
- A phase-locked loop (PLL) is an electronic system that generates a periodical signal with a frequency that is equal to a particular input signal frequency component.
- With a PLL, signals with frequencies equal to the sum or a multiple of two other frequencies can be generated.
- The important PLL parameters, defined as the allowable frequency range of the input signal before and after lock-in, are the hold range and the lock-in range.

- A lock-in amplifier is an AC amplifier based on synchronous detection. It makes it possible to measure the amplitude and phase of very small, noisy AC signals.
- A chopper amplifier is a DC amplifier in which the input signal is first modulated, than amplified as an AC signal and finally demodulated by synchronous detection. Such amplifiers have extremely low offset voltage and bias currents.

EXERCISES

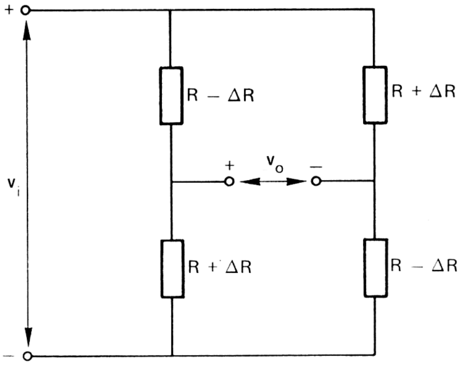
Amplitude modulation and demodulation

17.1 What is an amplitude modulated signal with a suppressed carrier? What is pulse width modulation?

17.2 If the amplitude of a 5 kHz frequency sine-wave carrier is modulated by a symmetric triangular signal where the fundamental frequency is 100 Hz what are all the frequencies in the modulated signal?

17.3 A frequency-multiplex system based on AM should transmit 12 measurement signals, each of which has a frequency band of 100-500 Hz. What, then, is the minimum bandwidth for this system?

17.4. In the Wheatstone bridge below all the resistors are sensors. The bridge is supplied with an AC voltage that has an amplitude of 10 V. The output v_o is amplified by a factor of 10^4 and multiplied by a synchronous signal with an amplitude of 4 V. Calculate the DC output voltage for a relative resistance change $\Delta R/R$ of 10^{-5} and -10^{-5}.



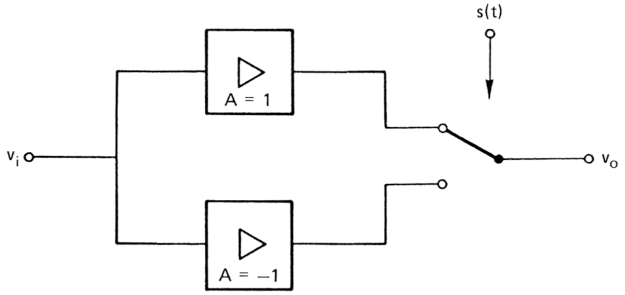
17.5 The question is the same as that posed in 17.4 with the difference that time it is a switched multiplier that demodulates the amplified bridge output signal in accordance with the following Fourier series $s(t) = \frac{4}{\pi}\{\cos\omega t + \frac{1}{3}\cos 3\omega t + + \frac{1}{5}\cos 5\omega t + ...\}$ and the switching signal is synchronous with the bridge power signal.

Systems based on synchronous detection

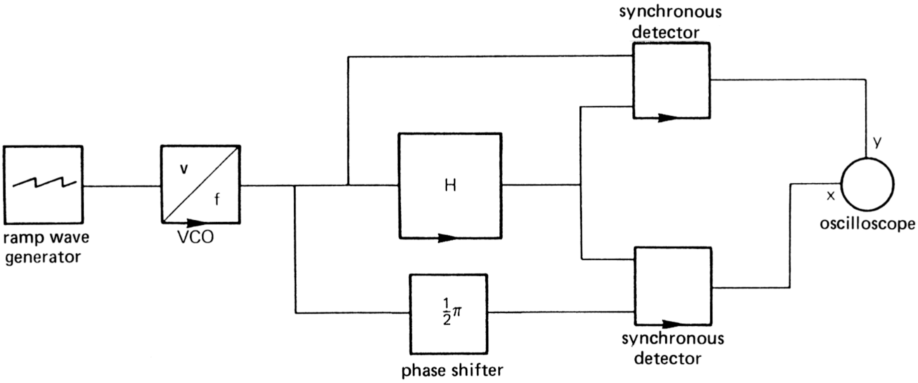
17.6 Imagine a synchronous detector composed of an ideal analog multiplier and a low-pass filter with a cut-off frequency of 100 Hz. The frequency of the reference signal is 15 kHz. What, then, are the input frequencies to which this detector is sensitive? Plot the sensitivity frequency characteristic.

17.7 The synchronous detector given in Exercise 17.6 actually acts as a band-pass filter. What is the detector's equivalent *Q*-factor?

17.8 You are given a synchronous detector composed of a switched modulator in line with the principle depicted below and a low-pass filter with a cut-off frequency of 200 Hz. The switching period is 0.2 ms. What are the frequencies detected by this system?



17.9 The figure below shows a block diagram of a system used to test a linear signal processing circuit. What is the characteristic that is displayed?



17.10 The Figure 17.12 comparator output is either high or low in accordance with the polarity of the phase shifter output. The output levels have accurate values. What are the requirements for the amplitude of the reference voltage v_{ref}, and how do they change when the comparator is omitted?

17.11 Sketch the frequency spectra of all the signals in the lock-in amplifier shown in Figure 17.12. Assume that the AC input signal is narrow-band.

17.12 Plot the frequency spectra of all the signals in the chopper amplifier given in Figure 17.14. Assume that the quasi DC input signal is narrow-band.

18 Digital-to-analogue and analogue-to-digital conversion

An analog signal is not suitable for digital processor or for computer processing. Only after having been converted into a digital signal can it be computer processed. Conversely, many actuators and other output devices require analog signals, so computer digital signals have to be converted into analog signals. Analog-to-digital converters (AD-converters or ADCs) and digital-to-analog converters (DA-converters or DACs) are available as modules, as integrated circuits and on data-acquisition cards for PCs. In the first part of this chapter a brief description will be given of digital signals and binary codes before going on to describe the main types of AD and DA-converters. The second part of the chapter will be devoted to certain particular types of converters.

18.1 Parallel converters

18.1.1 Binary signals and codes

A digital signal has just two levels, denoted as "0" and "1". The relationship with voltage values or current values will depend on the components implemented and, in particular, on the technology (such as: bipolar transistors, MOSFETs and possibly other kinds of transistors). With digital circuits that have, for instance, bipolar transistors a "0" corresponds to a voltage below 0.8 V while a "1" is represented by a voltage above 2 V. It is because of these wide tolerances that digital signals have much lower interference sensitivity than analog signals. This interference and noise tolerance leads, in turn, to a high reduction in information content. The minimum amount of information (yes or no, high or low, "0" or "1", on or off) is called a bit which is an acronym for “binary digit”.

Usually a measurement signal contains much more information than just 1 bit. If such information is to be adequately represented in binary signal form, whole groups of bits will be required, such groups are called binary words. A word consisting of 8 bits is called a byte (an impure acronym for “by eight”). The prefix “kilo (k)” stands for the decimal multiple 1000, so 1 kbyte = 1000 bytes. To express multiples of powers of two, the prefix Kibi (Ki) should be used, which equals $2^{10} = 1024$, so 1 Kibyte = 1024 bytes. The words byte and bit must not be confused but the notations kb and kB are unclear. In this book we will not therefore use those abbreviations, instead we shall simply refer to a kbyte or a kbit.

With n bits, just 2^n different words can be constructed. The number of bits is limited to a maximum, not only for practical reasons but also because of imperfections in the AD-converter components that generate the binary words. Analog-to-digital conversion thus causes at least one extra error, the quantization error (see also Section 2.1). Typical AD-converter word lengths range from 8 to 16 bits and correspond to a quantization error of 2^{-8} to 2^{-16}. Obviously it is useless to take an AD-converter with many more bits than what corresponds to the inaccuracy or the resolution of the input signal itself.

> ***Example 18.1***
> *The measurement signal range is 0 - 10 V. There are 10 bits available to represent this signal. The resolution of this representation is* $2^{-10} \approx 0.1\%$ *or about 10 mV.*
> *One other measurement signal has an inaccuracy of 0.01%. The number of bits required for proper representation is 14 because* $2^{14} > 10^4 > 2^{13}$.

A binary word can be written as:

$$G_2 = \left(a_n a_{n-1} ... a_2 a_1 a_0 a_{-1} a_{-2} ... a_{-m}\right) \tag{18.1}$$

where a_i is either 0 or 1 (in numbers). The value of G_{10} in the decimal number system is:

$$G_{10} = 2^n a_n + 2^{n-1} a_{n-1} + ... + 2^2 a_2 + 2a_1 + a_0 + 2^{-1} a_{-1} + 2^{-2} a_{-2} + ... + 2^{-m} a_{-m} \tag{18.2}$$

The coefficient a_n contributes most to G and is therefore called the most significant bit or MSB. The coefficient a_{-m} has the lowest weight and is therefore called the least significant bit or LSB.

The digital signals of AD and DA-converters are the binary coded fractions of a reference voltage V_{ref}. The analog voltage is equal to $V_a = G \cdot V_{re}$ where G is a binary number between 0 and 1. The relationship between the analog and digital converter signals is therefore:

$$V_a = V_{ref}\left(a_{n-1}2^{-1} + a_{n-2}2^{-2} + ... + a_1 2^{-n+1} + a_0 2^{-n}\right) = V_{ref} \sum_{i=0}^{n-1} a_i 2^{i-n} \tag{18.3}$$

Consequently the MSB of a converter corresponds to a value of $\frac{1}{2} V_{ref}$, the next bit is $\frac{1}{4} V_{ref}$ and so on until finally one arrives at the LSB with a value of $2^{-n} V_{ref}$.

Since very many bits are required to represent high numbers the binary notation system might be termed rather inefficient. This is why hexadecimal notation, based on the hexadecimal number system (base 16) has become common. The 16 digits are denoted as 1, 2, 3, ... 9, A, B, C, D and F. The last one, F, has a decimal value of 15 or a binary value of 1111. Hexadecimal notation is derived from binary notation by splitting up the latter into groups of four bits, starting from the binary point. The hexadecimal character is assigned to each group of four.

> ***Example 18.2***
> *The binary number 1010011110 can be written as 0010 1001 1110, or as 29E hexadecimal. In the decimal number system it is:*
>
> $2 \cdot 16^2 + 9 \cdot 16^1 + 14 \cdot 16^0 = 670.$
>
> *This can also be found from the binary notation:*
>
> $1 \cdot 2^9 + 0 \cdot 2^8 + 1 \cdot 2^7 + 0 \cdot 2^6 + 0 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 670.$

Other codes that are sometimes alternatively employed are the BCD code and the octal code. The BCD code (binary coded digit) is structured as follows:

$$\ldots\ a_2 b_2 c_2 d_2\ a_1 b_1 c_1 d_1\ a_0 b_0 c_0 d_0,$$

in which each group of four bits represents the binary coded decimal digit. This code is easier to interpret that the binary code.

> ***Example 18.3***
> *The BCD notation of the decimal number 670 in Example 18.2 is 0110 0111 0000 which are the binary codes of the respective decimal digits 6, 7 and 0.*

The octal code which is sometimes used in computer programs abides by the symbols 0, 1, 2, 3, 4, 5, 6 and 7. The code 10_8 stands for 8_{10} (the index denotes the number system). Octal notation can be directly derived from binary notation by creating divisions into groups of 3 bits (starting with the LSB), just like when converting from binary to hexadecimal notation. The number 670_{10} from the preceding example becomes 1236_8 in octal notation.

Figure 18.1 gives several electronic representations of a binary word. Figures 18.1a and b are dynamic representations (with voltages or currents as time signals). In Figure 18.1a, the bits of a word are generated consecutively: it is a serial word. The consequence is that it takes relatively long for each measurement value, which grows in proportion to the number of bits, to emerge. In Figure 18.1b the bits are generated and transported simultaneously, there is one line for each bit. There are as many parallel lines as there are bits in a word: it is therefore a parallel word. The information is available in a single moment but more hardware is required (cables, connectors, components). Figure 18.1c illustrates a static representation of the same binary word, this time with a set of switches. A "0" corresponds to a switch that is off and a "1" to a switch that is on.

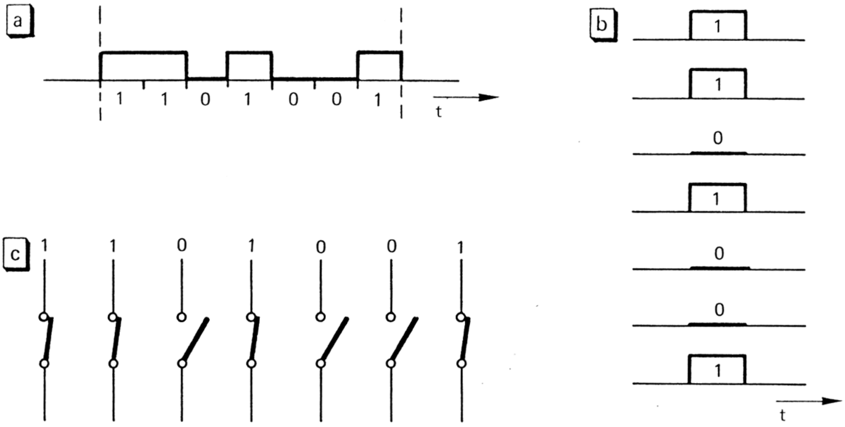


Figure 18.1. Various representations of a binary word (a) a serial word (dynamic), (b) a parallel word (dynamic), (c) a parallel word (static) using a set of switches.

The time required for one bit is called the bit-time. The transport of a serial n-bit word takes n bit-time, a parallel word can be transmitted substantially faster but it takes at least one bit-time. The conversion process for an analog signal $x(t)$ into a serial or parallel word takes time too, this is known as the conversion time. During conversion, the signal may change somewhat. To minimize the uncertainty of both the amplitude and the time of the converted signals, samples are taken at specific moments (see Section 1.2). Only the measurement values occurring at such moments are converted. The sampled signal should be fixed during the AD-converter conversion time. This is accomplished by having a sample-hold circuit (Section 15.2). If the conversion time is short compared to the actual changes in the input signal then an additional sample-hold circuit will not be required. The AD-converter itself performs the sampling.

18.1.2 Parallel DA-converters

This section only deals with one type of DA-converter, the parallel converter with a ladder network. It is the most widely used type when it comes to general applications and it is available at a very low price as an integrated circuit.

The first step in DA-conversion is to transfer from n parallel signal bits to sets of n parallel switches. In order to be able to activate the switches the binary signals have to satisfy certain conditions. Once the digital parallel input signal has been copied to the switches, the proper weighing factors must be assigned to each of the (equal) switches: half the reference voltage to the switch for the MSB (a_{n-1}), a quarter of the reference to the next switch a_{n-2}, and so on. Figure 18.2 exemplifies this assigning. A current is then allotted to each switch which corresponds to the weighing factor: $I/2$ for the MSB, $I/4$ for the next bit, and so on. The current for the last switch (LSB) is $I/2^n$. In this circuit, the switches have the two positions left and right (1 and 0, respectively). When in the right position the current flows directly to ground. In the left position it flows to a summing point. The sum of the currents is just the analog value that corresponds to the binary input code.

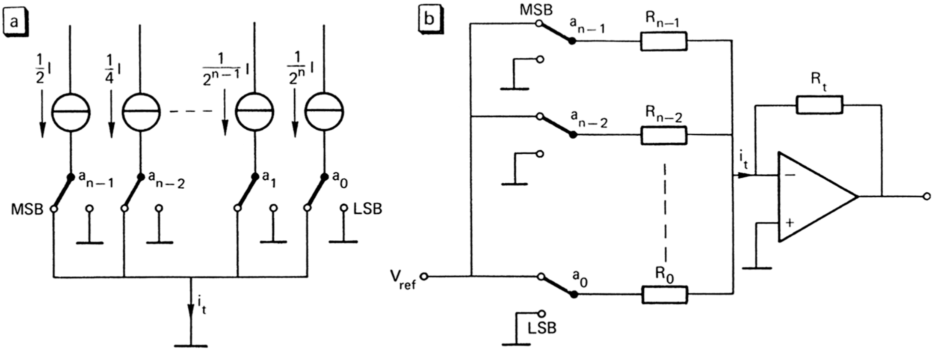


Figure 18.2. Converting a binary code into (a) a current, (b) a voltage.

The weighted currents $I/2$, $I/4$ etc. are derived from a single reference voltage (see Figure 18.2b). Since the non-inverting input terminal voltage for the operational amplifier is zero, the current through each resistor equals V_{ref}/R. The ratio between two subsequent resistance values is exactly 2 so that the corresponding currents also differ by a factor of 2. The total current then becomes:

$$i_t = V_{ref}\left(\frac{a_{n-1}}{R_{n-1}}+\frac{a_{n-2}}{R_{n-2}}+\ldots+\frac{a_1}{R_1}+\frac{a_o}{R_0}\right)=V_{ref}\sum_{i=0}^{n-1}\frac{a_i}{R_i}=V_{ref}\sum_{n=0}^{n-1}\frac{a_i 2^{-n+i+1}}{R_{n-1}} \qquad (18.4)$$

in which R_{n-1} the lowest resistance, belongs to the MSB a_{n-1}, and $R_i = R_{n-1}\cdot 2^{n-1-i}$. When the feedback resistor R_t is made equal to $R_{n-1}/2$, the output voltage of the converter is:

$$v_o = -i_t R_t = -V_{ref}\sum_{n=0}^{n-1} a_i 2^{-n+i} \qquad (18.5)$$

which, except for the minus sign, is equal to the required expression (18.1). V_{ref} is the full scale value of the converter: for all a_i equal to 1, v_o equals V_{ref} (minus 1 LSB).

The Figure 18.2b configuration is not often implemented because it contains both very high and very low resistances, in particular when there are large numbers of bits. It is rather difficult to create accurate resistors with a high LSB resistance value. Indeed, together with the unavoidable parallel capacitance, high resistance forms a large time constant, thus making the converter slow. On the other hand, very low resistances (for the MSB) may overload the reference source, thus resulting in unacceptable errors. These obstacles can be bypassed if the resistance network is replaced by the ladder network seen in Figure 18.3. One particular property of this network is the input resistance that is independent of the number of sections, as demonstrated by Figure 18.3.

Let us now consider the currents through this network. At the first (leftmost) node the input current I splits into two equal parts. One half flows through the resistor $2R$ while the other half goes towards the rest of the network as this also has an input resistance of $2R$. The latter current, $I/2$, splits up again at the second left node into two equal parts, and so on. The currents that pass through the $2R$ therefore have values $I/2$, $I/4$, $I/8$ etcetera. The network performs continual binary division using only two different resistance values, R and $2R$. This considerably simplifies the design of a DA-converter for large numbers of bits.

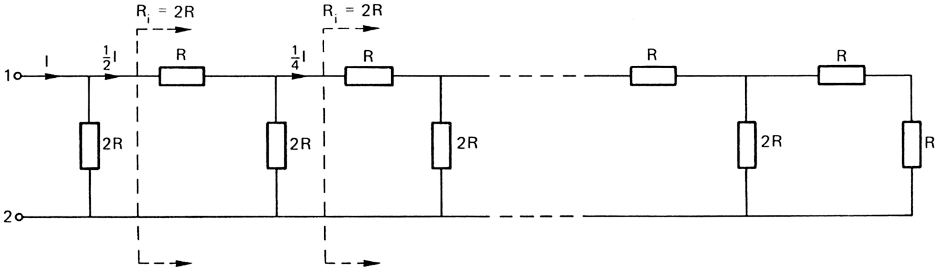


Figure 18.3. A ladder network containing two resistance values. The input resistance is independent of the number of sections.

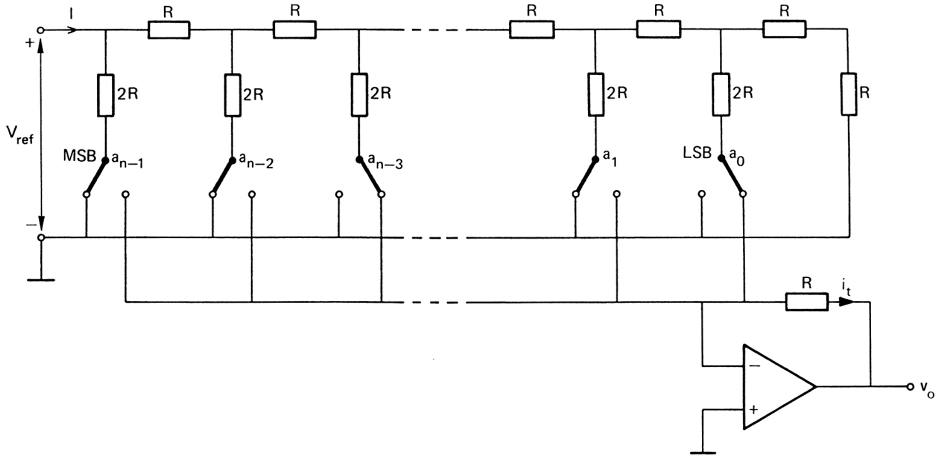


Figure 18.4. A digital-to-analog converter that is making use of a ladder network.

Figure 18.4 shows how the ladder network operates in a DA-converter circuit. The weighted currents either flow to ground (switch position "0") or to the summing point of the current-to-voltage converter (switch position "1"). The largest current $I/2$ flows to the switch for the MSB (a_{n-1}), the lowest current $I/2n$ to the switch for the LSB (a_0). The following equations are relevant to this circuit:

$$I = \frac{V_{ref}}{R} \tag{18.6}$$

and

$$i_t = \frac{a_{n-1}}{2} I + \frac{a_{n-2}}{2^2} I + \ldots + \frac{a_1}{2^{n-1}} I + \frac{a_o}{2^n} I = \sum_{i=0}^{n-1} \frac{a_i}{2^{n-1}} I \tag{18.7}$$

The output voltage is:

$$v_o = -i_t R = -R \sum_{i=0}^{n-1} \frac{a_i}{2^{n-1}} I = -V_{ref} \sum_{i=0}^{n-1} a_i 2^{-n+1} \tag{18.8}$$

which, again, is equivalent to Equation (18.3).
Table 18.1 lists the major specifications of an integrated 12-bit parallel DA-converter. This type of converter contains a built-in reference voltage (with Zener diode, Section 9) and an external connection that may be used for other purposes. The ladder network is composed of laser-trimmed SiCr resistors. The switches are made of bipolar transistors. In contrast to the circuit discussed above, this DA-converter has a current output. The smallest output current step is $2^{-12} \cdot 2$ mA which is about 0.5 μA. This current step corresponds to an input change of one binary unit. Such a 1-bit step is called an LSB as well. The inaccuracy and the other specifications of a DA-converter are usually expressed in terms of this unit, that is to say, the LSB (see Table 18.1). An inaccuracy of $\pm \frac{1}{2}$LSB corresponds (for this 12-bit converter) to $\pm \frac{1}{2} \cdot 2^{-12} \approx \pm 1.2 \cdot 10^{-4}$ of the full scale, or ±0.25 μA. The differential non-linearity represents the maximum deviation from the nominal smallest step (LSB) at the output. Guaranteed monotony means that the output never goes down when there is an upward change in the input code.

Table 18.1. The specifications for a 12 bit DA-converter (ppm = parts per million = 10^{-6}).

Input	'1': max. + 5.5 V, min. + 2.0 V '0': max. +0.8 V
Output current	unipolar –2 mA (all bits '1'), bipolar ±1 mA (all bits '1' or '0');
Output offset	< 0.05% full scale (unipolar)
Reference voltage	10 V ± 1 ppm/K (unipolar)
Inaccuracy	$\pm \frac{1}{4}$ LSB (25 °C), $\pm \frac{1}{2}$ LSB (0–70°C)
Differential non-linearity	$\pm \frac{1}{2}$ LSB (25°C)
Monotony	guaranteed (0–70°C)
Settling time	< 200 ns (up to $\pm \frac{1}{2}$ LSB)
Power dissipation	225 mW

When it is an output voltage that is desired rather than a current, the user must add a current-to-voltage converter to the DAC (Figure 18.5). An output voltage range of 0 – 10 V can be achieved by connecting the operational amplifier output to terminal 5 of the converter and by thus using the internal resistor $R = 5$ kΩ as a feedback resistor: $v_o = Ri_o$. The output voltage range can be doubled by making this connection to terminal 4 instead of to terminal 5 (see the dotted line in Figure 18.5). Other ranges can be made by connecting external resistances in series with internal resistances. By short-circuiting terminals 2 and 3, an extra current of 1 mA is added to the current-to-voltage converter. The input current of the operational amplifier will then run from −1 mA to +1 mA, thus resulting in a bipolar output voltage with a range of −5 to +5 V (or −10 to +10 V).

In the section above, the MSB is denoted as a_{n-1} and the LSB as a_0. Other notations may also be encountered, such as the reverse notation (a_0 for the MSB, a_{n-1} for the MSB), or from 0 (or 1) to n instead of to n−1: a_0 (or a_1) is the LSB, a_n is the MSB or vice versa).

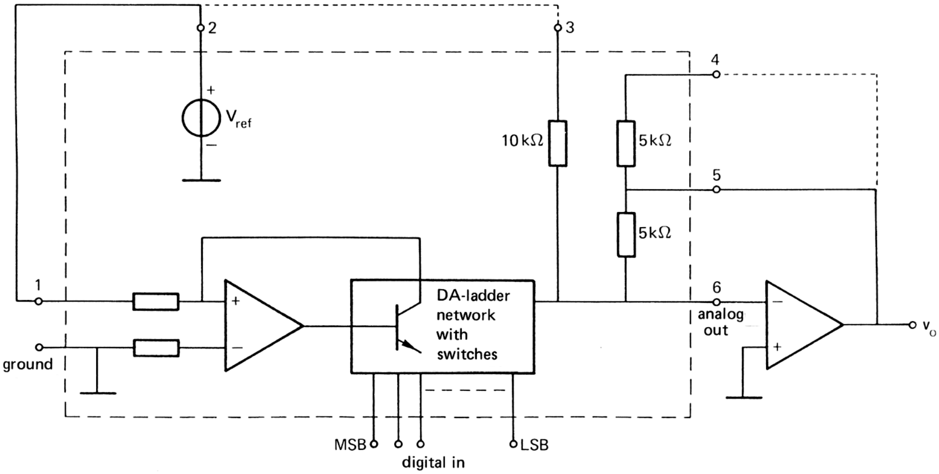


Figure 18.5. The internal structure of an integrated DA-converter with current output. A voltage output is obtained from an additional operational amplifier.

18.1.3 Parallel AD-converters

The output of an AD-converter is a binary code representing a fraction of the reference voltage, or current, corresponding to the analog input. Evidently this input signal must range between zero and the reference that is, it must cover the full scale of the converter. In this section we shall restrict ourselves to one type of AD-converter, the successive approximation AD-converter. This type of converter belongs to the class of compensating AD-converters that all use DA-converters in a feedback loop. The basic principle is illustrated in Figure 18.6.

Besides having a DA-converter, the AD-converter contains a comparator, a clock generator and a word generator. The word generator produces a binary code that is applied to the DA-converter input; this code also constitutes the AD-converter output. The word generator is controlled by the comparator (Section 14.2.1) whose output is "1" for $v_i > v_c$ and "0" for $v_i < v_c$, v_i is the AD-converter input and v_c is the compensation voltage which is identical to the DA-converter output. When the comparator output is "0", the word generator produces a following code corresponding to a higher compensation voltage v_c; when the output is "1", a new code is generated that corresponds to a lower value of v_c. This process continues until the compensation voltage is equal to the input voltage v_i (a difference of less than 1 LSB). As v_c is the output of the DA-converter, its input code is just a digital representation of the analog input voltage. In the final state, the average input of the comparator is zero, the input voltage v_i is compensated by the voltage v_c.

The clock generator produces a square wave voltage with fixed frequency (i.e. the clock frequency). It controls the word generator circuit so that at each positive or negative transition of the clock a new word is generated.

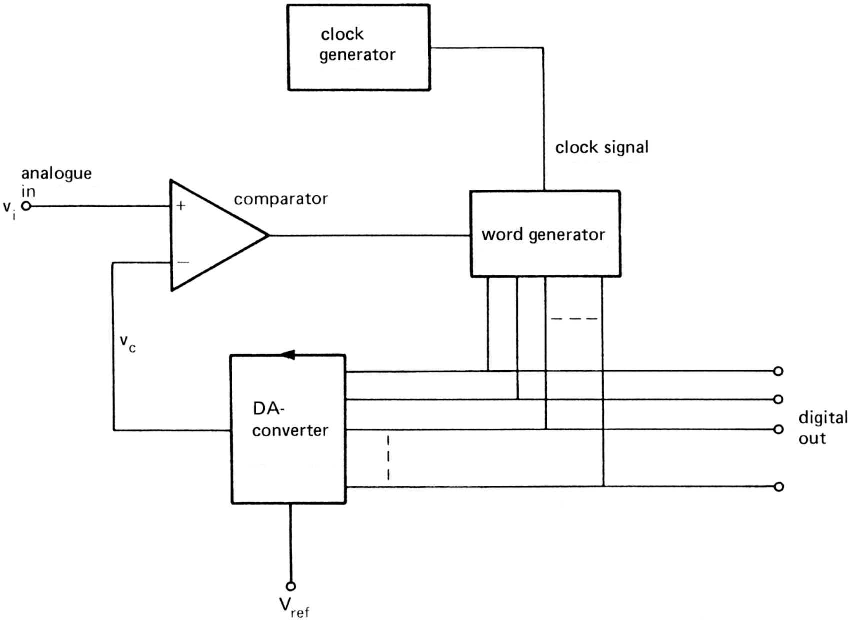


Figure 18.6. The principle of a compensating AD-converter using a DA-converter in a feedback loop.

The question which then arises is: how can the final state be reached as quickly as possible using the minimum number of clock pulses? As we do not know what the input signal is, the best estimation is halfway along the allowable range, that is $V_{ref}/2$. The first step in the conversion process is therefore to generate a code that corresponds to the compensation voltage of $V_{ref}/2$ (Figure 18.7). The comparator output will indicate whether the input voltage is in the upper half of the range ($V_{ref}/2 < v_i < V_{ref}$) or in the lower half ($0 < v_i < V_{ref}/2$). The comparator output is therefore exactly equal to the MSB a_{n-1} of the digital code we are looking for. If $a_{n-1} = 1$, then the compensation voltage will be increased by $V_{ref}/4$, in order to check during the next comparison whether the input is in the upper or the lower half of the remaining range. For $a_{n-1} = 0$, the compensation voltage is decreased by that same amount. The second comparison results in the generation of the second bit a_{n-2}. The successive comparisons are governed by the clock generator usually consisting of one comparison (or one bit) per clock pulse. After n comparisons have been made the LSB a_0 will be known and the conversion process finished.

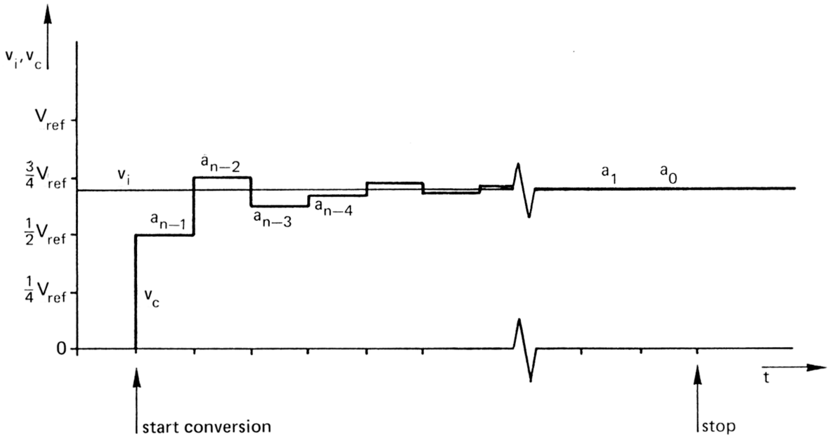


Figure 18.7. An amplitude-time diagram relating to the compensation voltage in a successive approximate AD-converter.

The word generator, which operates according to the principle explained here, is known as a successive approximation register or SAR. Most low-cost AD-converters are based on this principle. Though the bits are generated consecutively the bit values are stored in the digital memory (which is integrated into the converter) so that they remain available as parallel binary words. Some converter types have a serial output as well, where the bits are only available as series words during the conversion process.

The conversion time of a successive approximation AD-converter is equal to the number of bits n multiplied by the clock period T_c: $t_c = n \cdot T_c = n/f_c$ and independent of the analog input signal.

Table 18.2. The specifications of a successive approximate AD-converter.

Resolution	10 bit
Analog input	bipolar – 5 V up to + 5 V unipolar 0–10 V
Input impedance	5 kΩ
Offset (bipolar)	± 2 LSB (adjustable to zero) ± 44 ppm/K
Differential non-linearity	< ± $\frac{1}{2}$ LSB (25°C)
Conversion time	min. 15 µs; max. 40 µs
Power dissipation	800 mW

Table 18.2 reviews the main specifications of a 10-bit successive approximation ADC. This type of converter contains a clock generator (500 kHz), a reference voltage source, a comparator and buffer amplifiers at each output. Figure 18.8 shows the internal structure of this fully integrated converter. The functions of the terminals are as follows. The analog input voltage is connected between terminals 1 (ground) and 2 (note the relatively low input resistance). At floating terminal 3 the input voltage range goes from 0 to +10 V while at grounded terminal 3 an additional current source is connected to the internal DA-converter input. This prompts an input range shift from 0...10 V to –5...+5 V.

Terminals 4 to 13 form the 10-bit parallel digital output. Ten buffer amplifiers are connected between the DA-converter and the output terminals, these are called tri-state buffers. A tri-state buffer has three states: "0", "1" and off, in the off-state there is no connection between the converter and the output terminals. This third state is controlled by an extra input to each buffer. The tri-state buffers allow the complete circuit to be connected to one of the processor busses from which it can be electronically disconnected. This makes it possible for several devices to be connected to their corresponding terminals in parallel without the danger of short-circuiting.

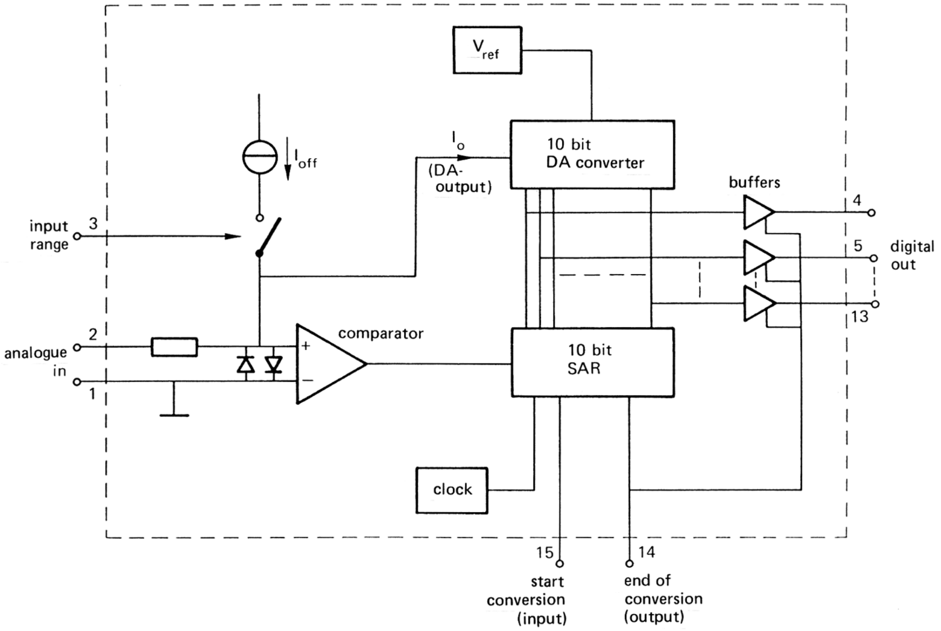


Figure 18.8. The internal structure of an integrated AD-converter with a successive approximation register (SAR).

As soon as the conversion process is finished the SAR generates a "0" at terminal 14 (during the conversion this output is "1"). The buffers connect the binary code to the corresponding output terminals. This output signal can furthermore be used to let the processor know that the conversion is finished and that the output data is valid (this terminal is also called the data ready output).

Finally, a binary signal at terminal 15 starts the converter. As long as this input is "1" the converter remains in a wait state but as soon as the input is switched to "0" (for instance by a signal from the computer), the conversion starts.

18.2 Special converters

In the first part of this chapter two popular types of AD and DA-converters were described. In this second part we shall introduce several other types: the serial DA-converter, two direct parallel AD-converters and the integrating AD-converter or dual slope converter.

18.2.1 The serial DA-converter

The weighing factors of a binary word, parallel or otherwise, follow from the position of the bits relative to the "binary" point. In an ADC or DAC these weighing factors are 1/2 (MSB), 1/4, 1/8 and so on up to $1/2n$ for the LSB. The allocation of the weighing factors is based on the spatial arrangement of the bit lines or the switches. In a serial word (Figure 18.9a) allocation is based on the time order of the bits. The first step is to convert the binary signal, which may be a voltage or a current, into a corresponding switch state (on or off). One side of the switch is connected to a reference voltage V_{ref} while the other terminal is zero when the switch is off and V_{ref} when the switch is on (Figure 18.9b). The conversion process takes place in a number of sequential phases. Let us now just suppose that the LSB is the bit in front. If that is the case, the first phase will consist of the following actions: a division of the voltage a_0V_{ref} by 2, and a storage of this value in a memory device. In the second phase, the value a_1V_{ref} will be added to the contents of the memory, the result will then be divided by 2 and stored again while the old value is deleted. The memory will then contain the value $(a_0/4)V_{ref}$ + $(a_1/2)V_{ref}$. This process will be repeated as long as bits arrive at the converter input. When the MSB (i.e. the nth bit) commences, the contents of the memory will be:

$$v_o = V_{ref}\left(a_0 2^{-n} + a_1 2^{-n+1} + \ldots + a_{n-2}2^{-2} + a_{n-1}2^{-1}\right) \tag{18.9}$$

which is precisely the desired analog value GV_{ref}. Figure 18.10 shows a circuit that works exactly in the way indicated by this procedure. The memory device is a capacitor which is charged when the switch S_2 is placed in the left-hand position. When the switch is in the right-hand position, the charge will be maintained and can be measured by a buffer amplifier. The sample-hold circuit retains the voltage when C is disconnected from the sample-hold circuit. Both switches, the bit switch S_1 and the sample-hold circuit switch, have equal switch rates that are also equal to the bit frequency. In each bit period S_2 samples half of the refreshed value and transfers that value to the sample-hold circuit.

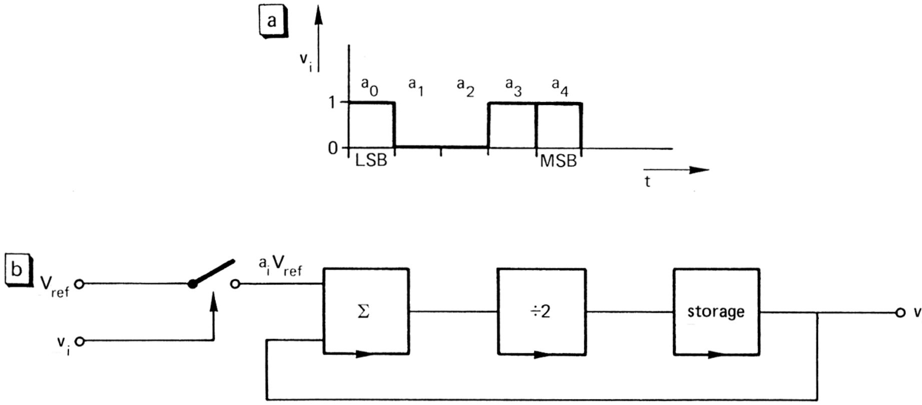


Figure 18.9. (a) An example of a binary serial word in which the LSB is the leading bit and (b) a functional diagram of a serial DA-converter.

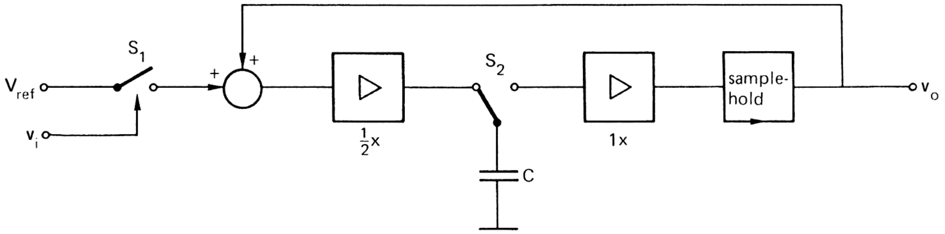


Figure 18.10. A DA-converter along the lines of the procedure given in Figure 18.9. The switches are controlled by a clock signal.

The structure of this converter does not depend on the number of bits. The inaccuracy is mainly down to the imperfections of both amplifiers and both memory devices (the capacitor C and the sample-hold circuit).

18.2.2 The direct AD converter

The structure of a direct AD-converter is quite straightforward (Figure 18.11). The input voltage is simultaneously compared to all possible binary fraction references. With an n-bit converter there are 2^n distinct levels. The reference voltage is subdivided into 2^n equal voltages. Using the same number of comparators the input voltage v_i is compared to each of these levels. There will be a certain number of comparator outputs, counting from the top which is low while the rest is high, and everything will depend on the input voltage. A digital decoder combines these 2^n values and generates the required binary code of n bits. This AD-converter is characterized by its high conversion speed limited by the comparator and decoder time delay as well as by the large number of components (and thus also the high price).

It is possible to significantly reduce the number of components but this will mean having to sacrifice speed (Figure 18.12). Here, the input voltage v_i is compared to $V_{ref}/2$, which results in the MSB. If $a_{n-1} = 1$ (thus $v_i > V_{ref}/2$), v_i will be reduced by $V_{ref}/2$ but otherwise remains the same. To determine the next bit, a_{n-2}, the corrected input voltage should be compared to $V_{ref}/4$. However, it is easier to compare twice the value with $V_{ref}/2$ which is the same. The voltage $v_i - a_{n-1}V_{ref}/2$ is therefore multiplied by 2 and compared to a second comparator with half the reference voltage. This procedure is repeated until the last bit (LSB) is reached.

Because the comparators are by this stage connected in series, this particular converter is called a cascaded DA-converter. It is somewhat slower than the preceding type, because the delay times accumulate, on the other hand the number of components required has been considerably reduced.

Table 18.3 shows a specification of the two DA-converters described above. Normally the conversion time is inverse to the conversion frequency. However, that no longer holds for very fast cascaded converters. The word bit frequency can be much higher than the conversion speed: a bit is applied to the previous comparator when the preceding bit is in the second comparator and so on (pipe-lining). This explains why the conversion frequency of the cascaded converter is higher than its inverse conversion time.

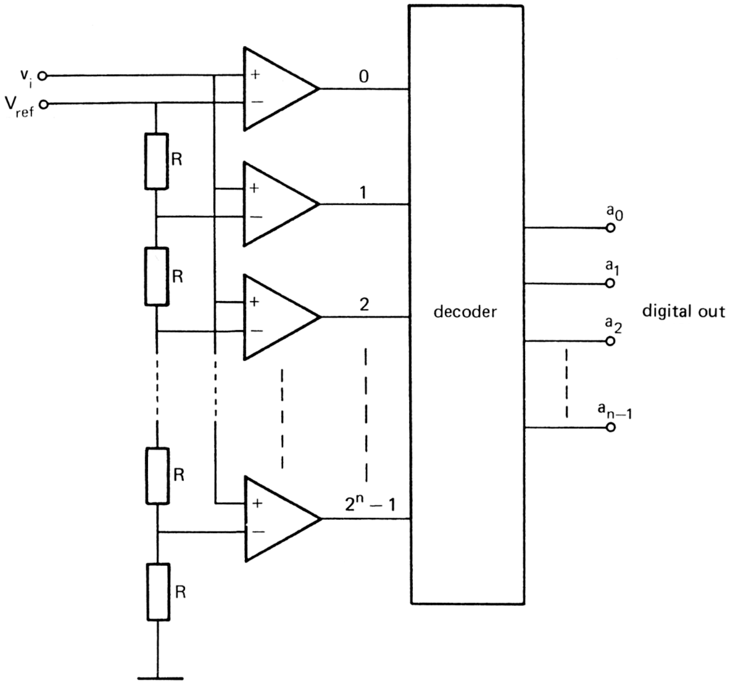


Figure 18.11. An *n*-bit direct AD-converter uses 2^n accurate resistors and 2^n comparators, the conversion speed is very high.

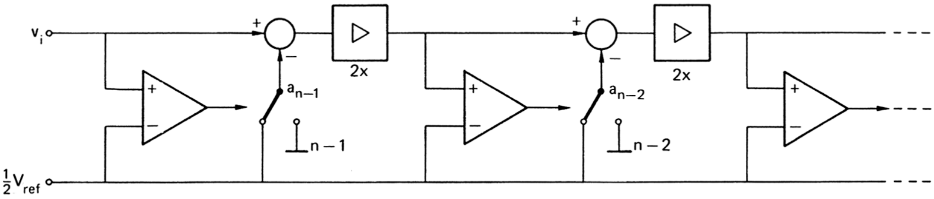


Figure 18.12. The principle of a cascaded AD-converter.

Table 18.3. A specification example for a direct and for a cascaded DA-converter.

	parallel	cascaded
resolution	8 bit	8 bit
inaccuracy	±0.15% ± ½ LSB	±0.15% ± ½ LSB
monotony	guaranteed	guaranteed
differential non-linearity	0.01%	0.01%
conversion time	35 ns	150 ns
conversion frequency	20 MHz	11 MHz

18.2.3 Integrating AD-converters

In an integrating AD-converter the input signal is integrated during conversion, the digital output is proportional to the mean value of the input signal. Since noise and other interference signals are also integrated, their contribution to the output is low (presuming that the average value of these signal components is zero). For this reason, integrating AD-converters are widely used in accurate DC current and voltage measurement systems. In connection with input integration these converters have a slower response than other AD-converters.

Figure 18.13 shows the principle of a dual-slope integrating AD-converter. First the input signal is integrated and then a reference voltage is given. The conversion starts with the switch S_1 in the upper position. The input signal is connected to the integrator which integrates v_i for a fixed time period T. When constant input is positive and there is positive integrator transfer, output will increase linearly in accordance with time. The comparator output is negative and keeps the switch, S_2, on. This switch lets past a series of pulses with a frequency of f_0 to a digital counter circuit. In digital voltmeters this is usually a decimal counter. At each pulse the counter is incremented by one count, as long as S_2 is on. When the counter is "full" (i.e. when it has reached its maximum value) it gives the command to switch S_1 to disconnect v_i and to connect the negative reference voltage to the integrator. At the same moment, the counter will again start counting from zero. Because V_{ref} is negative the integrator output decreases linearly in terms of time. As soon as the output detected by the comparator is zero, S_2 will switch off and the counter will stop counting. The content of the counter at that point in time constitutes a measure of the integrated input voltage. Figure 18.14 reveals the integrator output for one conversion period and two different constant input voltages.

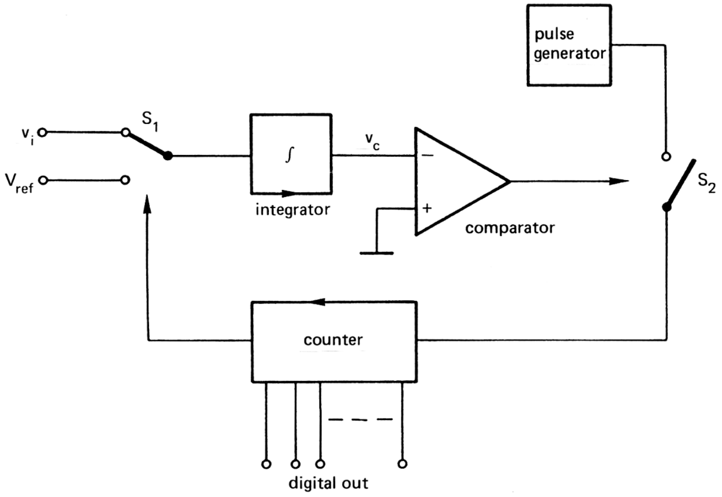


Figure 18.13. The principle of a dual slope AD-converter.

The procedure is as follows: the conversion starts at $t = t_0$, the counter is then reset to zero, switch S_2 goes on as soon as v_c is positive and the counter starts counting. By the time $t = t_0 + T$ the counter is full and by that stage the integrator output voltage is

$$v_c(t_0 + T) = \frac{1}{\tau} \int_{t_0}^{t_0+T} v_i dt \tag{18.10}$$

where τ is the proportionality factor (RC-time) of the integrator. At constant input voltage the output slope is v_i/τ. From $t = t_0 + T$ the negative reference is integrated, which means that v_c decreases at a rate of V_{ref}/τ. When $v_c = 0$ (at $t = t_1$), S_2 goes off and the counter stops. The voltage rise during the first integration (of v_i) is fully compensated by the voltage drop during the second integration (of V_{ref}). This maximum voltage is equal to the slope multiplied by the integration time so it is:

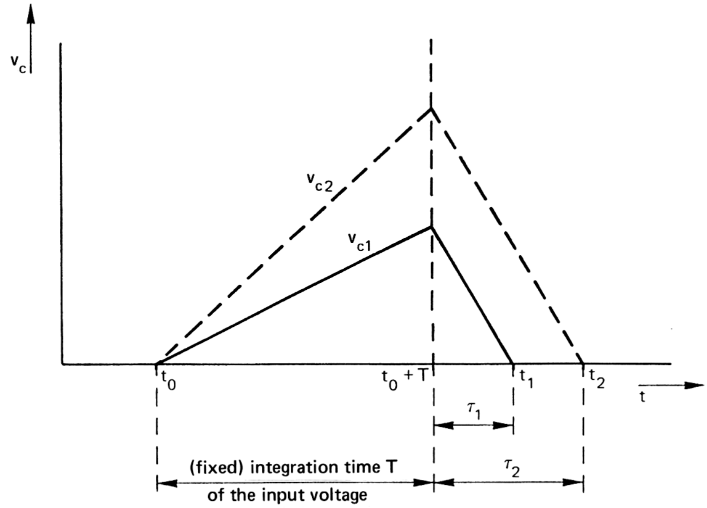


Figure 18.14. The output v_c of the integrator during one dual slope AD-converter conversion period for two different input voltages v_1 and v_2. The integration time for v_1 is τ_1; for v_2 it is τ_2 ($v_1 < v_2$). The integration time τ is proportional to the input voltage v_i.

$$\frac{v_i}{\tau} T = \frac{V_{ref}}{\tau} \left(t_1 - t_0 - T \right) \tag{18.11}$$

or:

$$v_i = V_{ref} \frac{t_1 - \left(t_0 + T \right)}{T} \tag{18.12}$$

The voltage v_i is transferred to a time ratio. This ratio is measured by the decimal counter. The first integration period T is such that the counter is full at a power of 10, so 10^n. When the counter starts again with the second integration, at $t = t_0 + T$, the number of pulses is equal to $(t_1 - t_0 - T) \cdot 10^n / T = (v_i / V_{ref}) \cdot 10^n$. If v_i is not constant during the integration period, the number of pulses N at the end of the second integration period equals:

$$N = \frac{10^n}{V_{ref}} \int_{t_0}^{t_0 + T} v_i dt \tag{18.13}$$

At first sight this conversion method might seem rather laborious. However, the resultant output code depends entirely on the reference voltage and, of course, on the input voltage but not on other component values. The only requirement placed upon the integrator and the frequency f_0 is that they be constant during the integration period.

The method also makes it possible to greatly reduce mains interference (from 50 or 60 Hz spurious signals). The integration time chosen to suppress 50 Hz interference is a multiple of 20 ms. In such cases the average for a 50 Hz signal is just zero.

Digital voltmeters based on the dual slope technique have an inaccuracy of less than 10^{-4} to 10^{-5}. The converters are available as module or integrated circuits.

SUMMARY

Parallel converters

- A binary signal has just two levels denoted as "0" and "1". The information load is restricted to 1 bit at a time, a binary signal is very fault tolerant.
- A binary word is a group of bits. A byte is a group of 8 bits and a kilobyte contains 1024 bytes. The bit with the highest weight is the most significant bit or the MSB while the bit with the lowest weight is the least significant bit or LSB.
- The ultimate number of bits limits the resolution of the measurement quantity, the error due to AD conversion is at least the quantization error: $\pm \frac{1}{2}$LSB.
- Besides the decimal and the binary number systems other common systems are the hexadecimal and octal systems (with bases 16 and 8 respectively). The BCD code is a combination of the decimal and the binary code.
- Binary words materialize dynamically (through voltages or currents) or statically (by means of switches) either as parallel words or serial words. A serial word can be transmitted along a single line but a parallel word needs as many lines as there are bits.
- Conversion from analog to digital signal form requires sampling and quantization. Both steps can lead to the introduction of additional errors.
- The digital signal of DA and AD-converters is the binary code of a fraction G of the reference (voltage): $V_a = GV_{ref}$, with $0 < G < 1$. The reference is the full scale. The weight of the MSB is just $\frac{1}{2}V_{ref}$.
- The inaccuracy parameters of AD and DA-converters are expressed in terms of LSB units. The LSB of an n-bit converter with reference voltage V_{ref} is $V_{ref}/2^n$.
- The ladder network in a parallel DA-converter generates a series of currents which differ successively by a factor of 2, they are the weighing factors of the bits.
- The differential non-linearity of an AD or DA-converter is the maximum deviation from the nominal step of 1 LSB. If differential non-linearity is more than 1 LSB monotony can no longer be guaranteed.
- Compensating AD-converters make use of a DA-converter in a feedback loop. A widely used type is the successive approximation converter. With n bits conversion is performed within n steps (involving comparisons and clock pulses).

Special converters

- A serial DA-converter converts a binary coded signal directly upon receipt of the bits using a capacitor as a memory device and a sample-hold circuit.
- An n-bit direct AD-converter has 2^n comparators but it is very fast. The cascaded converter is also fast but it has a reduced number of components (n comparators in series).
- An integrating AD converter responds to the integral or average of the analog input signal. These converters are slow but accurate because of the high noise and interference immunity.
- The two-fold integration (one the input, one the reference) of the dual slope integrator makes the system insensitive to component tolerances. It is therefore widely used in accurate voltage and current measurement systems.

EXERCISES

Parallel converters

18.1 Complete the next table.

binary	1010111						
octal		577					111
decimal			257			111	
hexadecimal				8F	111		

18.2 The reference voltage of a 10 bit DA-converter is 10 V. Calculate the output voltage when the input code is 1111100000 (MSB first).

18.3 The reference voltage of a 12 bit DA-converter has a temperature coefficient of ±2 ppm/K. Find the inaccuracy in the output voltage over a temperature range of 0 to 80 °C, expressed in terms of the LSB.

18.4 What is the differential non-linearity of a DA-converter? What is monotony?

18.5 Integrated digital circuits employ the binary number system and not, for instance, a number system with the base 4, where one value gives a choice of 1 out of 4. What is the reason for using this rather inefficient number system?

18.6 The clock frequency of a 10-bit successive approximation AD-converter is 200 kHz. Find the (approximated) conversion time for this converter.

18.7 Explain the term "multiplying DAC" for a DA-converter with external reference.

18.8 What is the function of the two diodes connected in anti-parallel at the input of the integrated circuit given in Figure 18.8?

Special converters

18.9 A serial binary signal with a bit frequency of 1 Mbit/s is applied to the serial DAC shown in Figure 18.10. Find the conversion time for a 14-bit serial word.

18.10 The input signal of the DAC in Figure 18.10 is the 3-bit word 101. Make a plot of the relevant output signal versus time. The capacitor is uncharged for $t < 0$.

18.11 The specifications of the 3-bit cascaded converter seen in Figure 18.12 are:

- reference voltage $V_{ref} = 5.000$ V;
- offset voltage of the comparators: ±6 mV;
- offset voltage of the amplifiers: ±6 mV;
- inaccuracy of the gain factor: ±0.5 %;

Calculate the maximum digital output error caused by each of these specifications expressed in LSB for an input voltage of 0.630 V.

18.12 Explain why the pulse frequency is not of importance to the dual slope converter.

18.13 The integration period of an integrating AD-converter is 100 ms ±1 μs. Determine the maximum conversion error caused by a 50 Hz interference signal with an rms value of 1 V.

19 Digital electronics

What we shall be looking at in this chapter is the kind of integrated circuits that are used to process digital signals. Before going on to provide a functional description of certain digital components, Section 19.1 starts by introducing Boolean algebra. Section 19.2 then goes on to deal with several widely-used types of digital circuits such as: multiplexers, adders, counters and shift registers.

19.1 Digital components

19.1.1 Boolean algebra

Analog signals and analog transfer functions can be adequately described with the aid of time functions and frequency spectra. Such descriptions are useless when applied to digital signals and digital processing circuits. For such purposes we employ a certain mathematical method which was first used to describe logical processes and which was devised by George Boole in 1847.

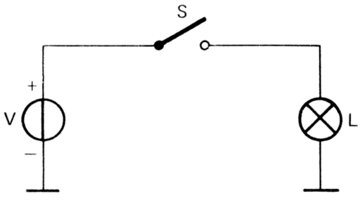


Figure 19.1. If the statement "the switch is on" is true, then the statement "the lamp is on" must also be true.

The two values underpinning logical statements are: "true" (T) and "false" (F). Taking various examples from the field of electronic application we shall endeavor to illustrate this concept. Figure 19.1 shows a lamp and switch circuit. In conjunction with this particular system there are a number of assertions that can be made like, for instance: l "the lamp is on" and s "the switch is on". Both statements can be either true or false but if l is true then s has to be true as well. The relations between various statements can be described by means of, for instance, a truth table. In relation to the statements made above Table 19.1 is one such truth table for the circuit shown in Figure 19.1.

Table 19.1. Truth table for the circuit given in Figure 19.1. s = "the switch is on"; ℓ = "the lamp is on".

s	ℓ
false	false
true	true

Table 19.2. Truth table for the circuit seen in Figure 19.1. s = "the switch is on"; ℓ = "the lamp is off".

s	ℓ
false	true
true	false

Obviously other statements can be defined and the truth tables will change accordingly but the physical operations will not, of course, change. Table 19.2 is another truth table for the same circuit shown in Figure 19.1 but accompanied by other statements.

Let us just suppose that we have a lamp circuit with two switches in series (Figure 19.2). If that is the case then the truth table will be extended to provide four possible combinations for the two switches. Table 19.3 provides a truth table for the same statements as those given before but this time they are denoted as a and b: "the switch is on" and l: "the lamp is on". For the sake of simplicity we shall adopt the symbols T and F for true and false. As can be expected from the figure, the lamp can only be said to be on (l = T) when both switches are on (a = b = T). This is an example of how the logic AND works : $l = a$ AND b; the statement l is only true if both a and b statements are true. There are several notations for this operation: AND, $\wedge$, or $\cdot$, in this book we use the symbol $\cdot$, or it is left out $l = a{\cdot}b = ab$.

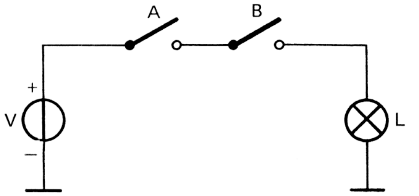


Figure 19.2. A circuit with two switches in series is described by the way the logic AND operates.

Table 19.3. The truth table for the circuit in Figure 19.2.

a	b	ℓ
F	F	F
F	T	F
T	F	F
T	T	T

Another way to control a lamp with two switches is illustrated in Figure 19.3 and the corresponding truth table is given in Table 19.4. In such a case at least one of the two switches has to be on to light the lamp. There is a correspondence between this operation and the logic operation OR: $l = a$ OR b. Possible notations are OR, $\vee$ or +. In this book we abide by the + symbol. From the truth table it follows that this is an "inclusive OR": either a or b or else a and b allow the lamp to ignite and this third possibility is included.

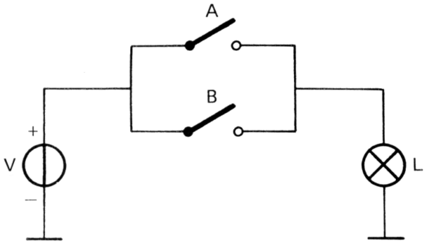


Figure 19.3. The OR logic operation describes a circuit with two switches in parallel.

Table 19.4. The truth table for the Figure 19.3 circuit. a = b = "the switch is on"; ℓ = "the lamp is on".

a	b	ℓ
F	F	F
F	T	T
T	F	T
T	T	T

The "exclusive OR" (or EXOR) is an operation as characterized in Table 19.5, only one of the statements *a* and *b* needs to be true to allow *l* to be true. This operation is denoted as EX-OR, EXOR, ⊕. In this book we use the symbol ⊕ which means that $l = a \oplus b$. The physical realization of the exclusive OR operation is given in Figure 19.4 with the two-way switch.

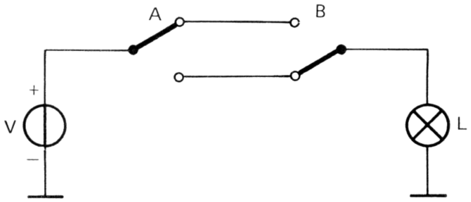


Figure 19.4. A two-way switch explained by means of an EXOR operation.

Table 19.5. The exclusive OR operation truth table.

a	b	ℓ
F	F	F
F	T	T
T	F	T
T	T	F

The statements given above cannot be applied to these types of switches. Instead we need to claim "the switch is in its rest position" thus meaning the position as shown in the figure.

A fourth logical operation is inversion or negation, denoted as NOT *a* or $\overline{a}$, we always use the latter notation. If *a* is a statement, for instance "the lamp is on" and if *a* is indeed true then the lamp is on. The lamp is also on when $\overline{a}$ (NOT *a*) is false. The inversion truth table is given in Table 19.6.

In digital electronics the alternative symbols used for F and T are "0" and "1". Note that the symbols "0" and "1" simply represent the variables false and true and should not be confused with the binary numbers 0 and 1. The symbol choice is, in principle, arbitrary. Physically, the logic variables can be presented in various ways, for instance to denote low and high voltage, current or no current, high or low frequency and so on. In digital circuits, a "0" is usually a voltage below 0.8 V and a "1" is usually a voltage above 2 V. Data transmission (i.e. telephone line systems) use two different frequencies or phases.

From now on we shall use the symbols 0 and 1 without quotation marks. Note again that there is a difference between the logic variables 0 and 1 (which are actually "0" and "1") and the numbers 0 and 1. Table 19.7 shows the basic logic operations in a truth table but this time with the symbols 0 and 1. The symbols a, b and l represent logic variables and have only two values: 0 or 1. In composite logic equations, like $z = x + y\overline{w}$, the logic operations order becomes: inversion - logic AND - logic OR. When the order is changed this is indicated by brackets just as in normal algebraic equations: $z = (x + y)\overline{w}$. The Boolean algebra has a number of rules to facilitate logic variable calculations. These rules can be proved by drawing up a complete truth table for all the possible combinations bearing in mind that the left and right-hand sides of the logic equations must be equal.

Table 19.6. The truth table for logic negation or inversion.

a	$\overline{a}$
F	T
T	F

Table 19.7. The truth table for $a{\cdot}b$, $a{+}b$, $a{\oplus}b$, $\overline{a}$ and $\overline{b}$.

a	b	$a{\cdot}b$	$a{+}b$	$a \oplus b$	$\overline{a}$	$\overline{b}$
0	0	0	0	0	1	1
0	1	0	1	1	1	0
1	0	0	1	1	0	1
1	1	1	1	0	0	0

* The following rules are valid for the logic values 0 and 1:

$$\begin{array}{lll} 0\cdot 0=0 & 0+0=0 & \\ 0\cdot 1=0 & 0+1=0 & \overline{0}=1 \\ 1\cdot 0=0 & 1+0=1 & \overline{1}=0 \\ 1\cdot 1=1 & 1+1=1 & \end{array} \tag{19.1}$$

Below is a list of various rules for general logic variables.

* **law of equality**:

$$\begin{array}{l} a\cdot a\cdot a\cdot\ldots=a \\ a+a+a+\ldots=a \end{array} \tag{19.2}$$

* **commutative laws for addition and multiplication**:

$$\begin{array}{l} a\cdot b=b\cdot a \\ a+b=b+a \end{array} \tag{19.3}$$

* **associative laws for addition and multiplication**:

$$\begin{aligned}&(a\cdot b)\cdot c = a\cdot(b\cdot c)\\&(a+b)+c = a+(b+c)\end{aligned}\tag{19.4}$$

* **distributive laws**:

$$\begin{aligned}&a\cdot(b+c) = ab+ac\\&a+(b\cdot c) = (a+b)(a+c)\end{aligned}\tag{19.5}$$

* **modulus laws**:

$$\begin{aligned}&0\cdot a = 0 \qquad 1\cdot a = a\\&0+a = a \qquad 1+a = 1\end{aligned}\tag{19.6}$$

* **negation laws**:

$$\begin{aligned}&a\cdot\overline{a} = 0\\&a+\overline{a} = 1\\&a = a\end{aligned}\tag{19.7}$$

$$\left.\begin{aligned}&\overline{a\cdot b} = \overline{a}+\overline{b}\\&\overline{a+b} = \overline{a}\cdot\overline{b}\end{aligned}\right\}\ \text{De Morgan's theorem}\tag{19.8}$$

* **absorption laws:**

$$\begin{aligned}&a\cdot(b+a) = a\\&a+a\cdot b = a\end{aligned}\tag{19.9}$$

$$\begin{aligned}&a\cdot(\overline{a}+b) = a\cdot b\\&a+\overline{a}\cdot b = a+b\end{aligned}\tag{19.10}$$

As can be seen from these formulas each logic equation has a counterpart known as the dual equation. This dual form is established by interchanging each 0 with 1 and vice versa and by interchanging each + by · and vice versa. The laws are used to simplify logic expressions and therefore also the logic circuits described by these equations.

Digital operations can also be made explicit with Venn diagrams. Some of these diagrams are given in Figure 19.5 to illustrate the operations $A\cdot B$, $A + B$, $\overline{A}$ and $A \oplus B$. This method is of no use when it comes to complex logic operations.

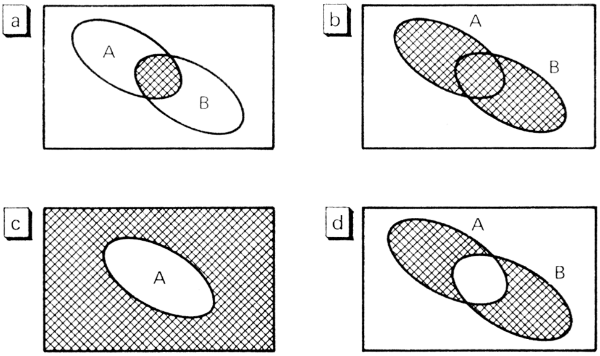


Figure 19.5. A logic operations representation using Venn diagrams: (a) $A \cdot B$, (b) $A+B$, (c) $\overline{A}$, (d) $A \oplus B$.

19.1.2 Digital components for combinatory operations

The two categories of digital circuits that exist are combinatory and sequential circuits. The output of any combinatory circuit is determined exclusively by the combination of actual input signals. The output of any sequential circuit can also depend on previous input values: a sequential circuit has memory properties.
Figure 19.6 provides an overview of the most commonly used elements of logic, their symbols and the corresponding Boolean equations. The American symbols are given as well as the official IEC symbols (International Electrotechnical Commission).

The elements given in Figure 19.6 are known as logic gates. They can have more than two inputs but only one output. Here below is a short description of the gates shown in Figure 19.6.

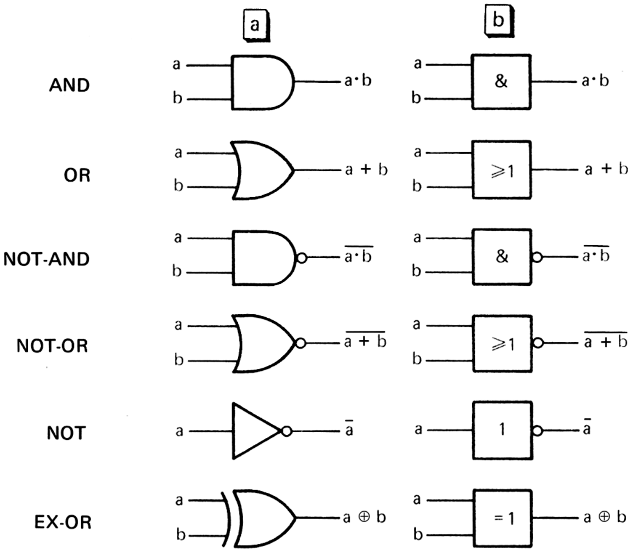


Figure 19.6. (a) The American and (b) the European logic element symbols.

- **AND-gate**

The output is 1 if all the inputs are 1. It functions in a similar way to a set of switches in series (Figure 19.2).

- **OR-gate**

The output is 1 if one or more inputs are 1. The OR-gate functions as a set of parallel switches (Figure 19.3).

- **Inverter (NOT)**

The output is the input complement. This element is usually combined with other gates. It is symbolized by a small circle at the input or output (see for instance the following NAND-gate).

- **NOT-AND or NAND**

The output is only 0 if all inputs are 1. It is an AND-gate in series with an inverter. The function is similar to a series of switches parallel to the load (Figure 19.7). The lamp remains on for as long as all the switches are not on.

- **EX-OR or EXOR**

This is an OR gate in the literal sense: the output is only 1 if either input *a*, input *b* (or input *c* etc.) is 1, in other words, if only one of the inputs is 1. The two-way switch in Figure 19.4 is an example of a circuit with an EXOR function.

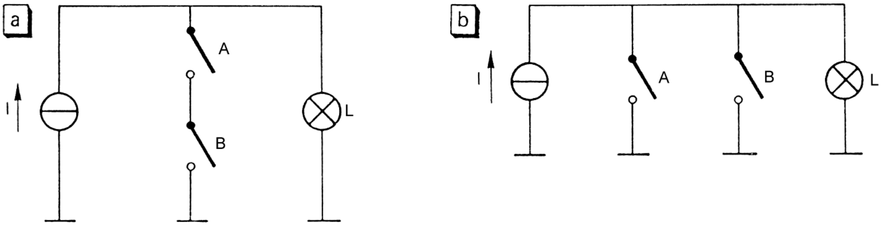


Figure 19.7. (a) A circuit with a NAND function, the dual form of the Figure 19.3 circuit, (b) a circuit with a NOR function, the dual form of the Figure 19.2 circuit.

Logic gates are available as integrated circuits. Usually, all depending on the number of pins, there is more than one gate in a single IC. For instance, an IC with eight inputs may comprise four gates with only two input terminals (is a "quad"), three gates each with three inputs (is a "triple"), two gates with 4 inputs (is a "dual") or one 8-input gate. Inverters are available in groups of six in one encapsulation (known as a "hex"). Figure 19.8 shows the internal structure of several digital ICs.

One can also have integrated circuits with a combination of gates for a particular function (adders, multiplexers). These circuits will be described in the second part of this chapter.

Gates are composed of active electronic components, like transistors. To operate properly they need to be powered. Most gate ICs require a voltage supply of 5 V. These are called TTL circuits (Transistor-Transistor Logic) and they are circuits that are composed of bipolar transistors, diodes and resistances. Figure 19.9a provides an example of a TTL NAND circuit.

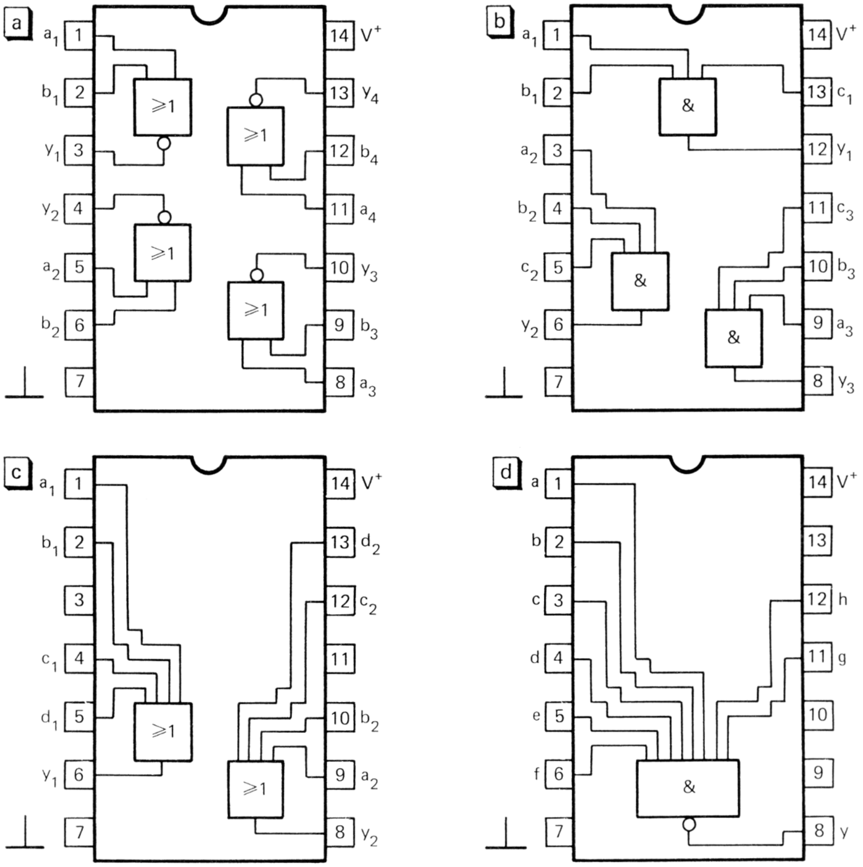

a
a1 1
b1 2
y1 3
y2 4
a2 5
b2 6
7
14 V+
13 y4
12 b4
11 a4
10 y3
9 b3
8 a3
≥1
b
a1 1
b1 2
a2 3
b2 4
c2 5
y2 6
7
14 V+
13 c1
12 y1
11 c3
10 b3
9 a3
8 y3
&
c
a1 1
b1 2
3
c1 4
d1 5
y1 6
7
14 V+
13 d2
12 c2
11
10 b2
9 a2
8 y2
≥1
d
a 1
b 2
c 3
d 4
e 5
f 6
7
14 V+
13
12 h
11 g
10
9
8 y
&

Figure 19.8. The internal structure of various logic gate ICs: (a) a quad 2-input NOR-gate, (b) a triple 3-input AND-gate, (c) a dual 4-input OR-gate, (d) an 8-input NAND-gate.

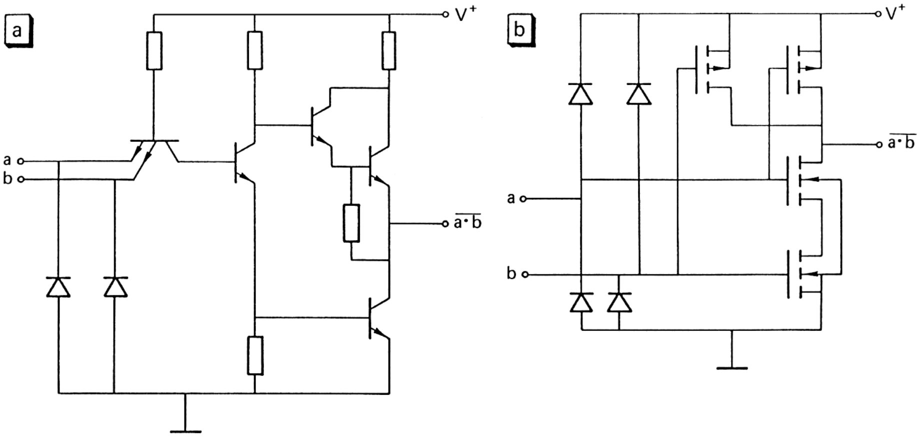

a
V+
a
b
$\overline{a \cdot b}$
b
V+
$\overline{a \cdot b}$
a
b

Figure 19.9. The internal structure of a NAND gate created (a) in bipolar TTL technology, (b) in CMOS technology.

Gates composed of MOSFETs (Section 11.1.2) also operate at a supply voltage of 5V. Components that are the result of CMOS technology (Complementary MOS, i.e. circuits with both p-channel and n-channel MOSFETs) operate at supply voltages of between 3 and 15 V. The particular properties of CMOS components are the very high input impedance ($\approx 10^{12}\ \Omega$) and the very low power dissipation. The power consumption of CMOS integrated circuits is extremely low, so much so that they can easily be battery powered. Batteries are discharged even quicker by their own leakage current than by the CMOS circuits. Figure 19.9b shows the internal structure of a CMOS NAND-gate. Precisely how it operates is not explained.

19.1.3 Digital components for sequential operations

Sequential circuits produce an output that not only depends on the actual input combination but also on previous combinations, in other words, they exhibit a memory function. The most important exponent in this category is the flip-flop. A flip-flop is a digital component with two or three inputs and two complementary outputs, it consists of a combination of gates. We shall only be describing two types of flip-flops, the SR flip-flop and the JK flip-flop.

19.1.4 The SR flip-flop

The simplest SR flip-flop is constructed with two NOR-gates (see Figure 19.10a). The element has two inputs, called the s (set) and r (reset) inputs. The two outputs are q and $\overline{q}$. When $sr = 10$ (which is short for $s = 1$ and $r = 0$), the output is 1 ($q = 1$, $\overline{q} = 0$). For $sr = 01$, the output is 0. When we start with one of these input combinations (10 or 01) and change to $sr = 00$, the output remains unchanged: the flip-flop recalls which of the inputs r or s was 1. This principle is illustrated in Table 19.8, the flip-flop truth table. The combination $sr = 11$ should be avoided: the two outputs are not complementary (both are 1). Furthermore, when changing from this state to $sr = 00$, the output will be either 0 or 1, depending on which gate has the fastest response and bearing in mind that the output state cannot be foreseen.
The symbol for the RS flip-flop is given in Figure 19.10b.

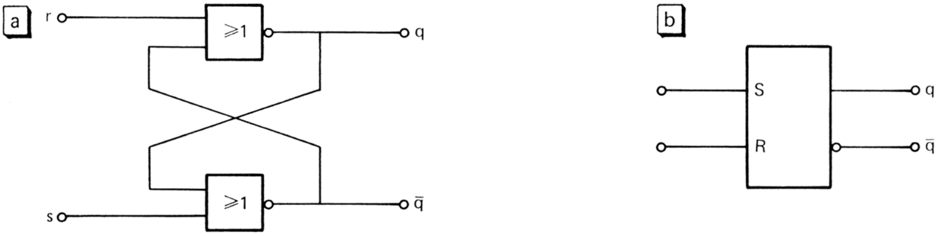


Figure 19.10. (a) A circuit configuration and (b) the symbol for an RS flip-flop.

Table 19.8. The SR flip-flop truth table of Figure 19.10. The symbol "-" means: 0 or 1, a "don't care", q_n denotes the output prior to a clock pulse, q_{n+1} indicates the output after that clock pulse.

s	r	q_n	q_{n+1}	$\bar{q}_{n+1}$	
0	0	–	q_n	$\bar{q}_n$	store
0	1	–	0	1	reset
1	0	–	1	0	set
1	1	–	0	0	not allowed
1→0	1→0	0	?	?	

The flip-flop output changes, if necessary, when there is a change in r or s. In circuits with many flip-flops it is often necessary to activate all the flip-flops at the same time. In such instances, the flip-flop will be extended by two AND-gates that have a common input, the clock input. The clock signal is a square wave voltage used to synchronize digital circuits. In such cases as these the clock controls the transferring of the input values r and s to the flip-flop inputs R and S. Figure 19.11a illustrates just how this works.

With a zero clock ($c = 0$), the outputs of the AND-gates are zero, irrespective of what r and s are and the flip-flop finds itself in the hold mode. As soon as $c = 1$, the outputs of the AND-gates become equal to R = r, S = s: the flip-flop is either set (rs = 01) or reset (rs = 10) (or else it remains unchanged when rs = 00). Figure 19.11b shows the circuit symbol for the clocked RS flip-flop in which C is the clock-input.

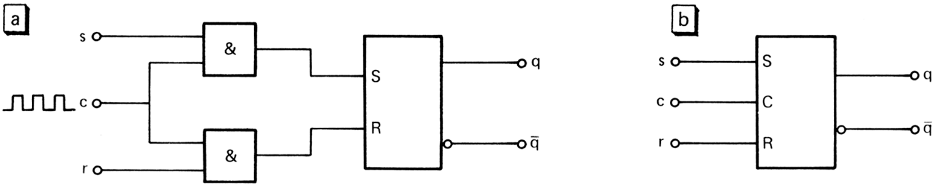


Figure 19.11. (a) An RS flip-flop with clock-input "c" that is used to synchronize several flip-flops in a circuit, (b) the circuit symbol for a clocked RS flip-flop.

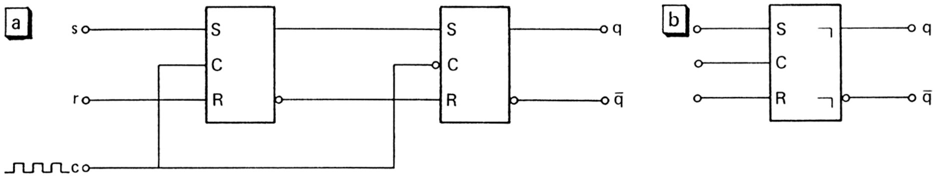


Figure 19.12. (a) A master-slave flip-flop composed of two clocked SR flip-flops. One flip-flop responds to the rising edge of the clock while the other responds to the clock's falling edge, (b) a circuit symbol for the master-slave flip-flop, distinguished by the ¬-sign.

When two RS flip-flops are connected in series (Figure 19.12a) binary information (i.e. 1 bit) is transferred from the first flip-flop to the second one and the clock serves as the command signal. In Figure 19.12a, the clock input for the second flip-flop is inverted (as is indicated by the small circle). The first flip-flop responds to the positive edge of the clock (from 0 to 1) and the second one responds when the clock changes from high to low. This combination is called a master-slave flip-flop arrangement. At $c = 1$, the first flip-flop (the master) transfers the information from the input to the output, the second flip-flop (the slave) stays in the hold mode. When there is a transition from $c = 1$ to $c = 0$ ($\bar{c} = 1$) the information is transferred to the slave output, while the master stays in the hold mode. A master-slave flip-flop therefore transfers the input to the output in two phases. The advantage of such a two-phase action is the time separation between the input and output changes thus allowing a direct connection to be made between flip-flop outputs and the inputs of other flip-flops. Time delay and transients have no impact on operations. The master-slave flip-flop symbol is given in Figure 19.12b.

19.1.5 JK flip-flops

The JK flip-flop is a master-slave flip-flop that is controlled by a clock signal. It is undoubtedly the most widely used of all flip-flops. The two inputs are called J and K while the outputs are q and $\bar{q}$. The four possible input combinations each have a different effect on output. The truth table (Table 19.9) shows the four modes of operation.

Table 19.9. Two types of JK flip-flop truth tables, the symbol "-" means 0 or 1, q_n is the output after the nth clock pulse and q_{n+1} the output after the (n+1)st clock.

j	k	q_{n+1}
0	0	q_n
0	1	0
1	0	1
1	1	$\bar{q}_n$

q_n	q_{n+1}	j	k
0	0	0	–
0	1	1	–
1	0	–	1
1	1	–	0

The truth table indicates how after the nth clock pulse the output q_n changes to q_{n+1} one clock pulse later. For $jk = 00$, the flip-flop is in the hold mode, for $jk = 10$, the flip-flop is set ($q = 1$ after the next clock pulse) and for $jk = 01$ the flip-flop is reset to $q = 0$. Up until this point, operation is similar to SR flip-flop operation. With the combination $jk = 11$ the output is inverted at the following clock pulse (it is said to "toggle"). In this mode, the flip-flop acts as a frequency divider: a clock signal with frequency f results in q, a square wave output that has half the clock frequency (Figure 19.13a).

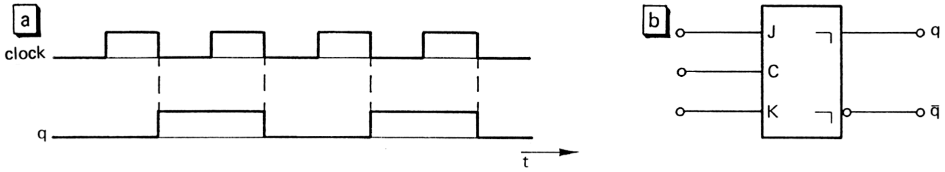


Figure 19.13. (a) A JK flip-flop can act as a digital frequency divider for $jk = 11$: the output frequency is half the clock frequency, (b) the circuit symbol of a master-slave JK flip-flop.

The circuit symbol for the master-slave JK flip-flop is given in Figure 19.13b.

Figure 19.14 shows the pin connection diagram that accompanies a commercial type of JK flip-flop. The integrated circuit contains two totally independent flip-flops in a 16-pin encapsulation. Apart from the J and K inputs the flip-flops in this IC have two other inputs: S (set or preset) and R (reset or clear). These inputs independently control the clock output and they are called asynchronous, in contrast to the J and K-inputs which are synchronous. The truth table (Table 19.10) shows the various modes for this flip-flop.

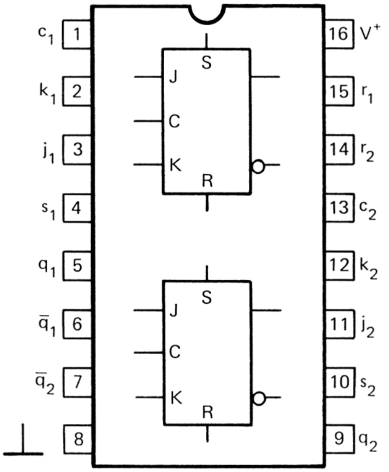


Figure 19.14. The internal structure of a dual JK flip-flop with asynchronous set and reset inputs.

Table 19.10. The truth table for a JK flip-flop with asynchronous set and reset possibilities. The sign ↓ means that the flip-flop triggers at the negative edge of the clock (from 1 to 0). "-" means 0 or 1 (it does not matter which), q_n is the output after the nth clock pulse, q_{n+1} is the output after the (n+1)st clock.

	Inputs			Outputs			
Modes	s	r	c	j	k	q_{n+1}	$\bar{q}_{n+1}$
asynchronous reset (clear)	0	1	–	–	–	0	1
asynchronous set	1	0	–	–	–	1	0
not allowed	1	1	–	–	–	1	1
hold mode	0	0	↓	0	1	0	$\bar{q}_n$
synchronous reset (0)	0	0	↓	0	1	0	1
synchronous set (1)	0	0	↓	1	0	1	0
synchronous inversion	0	0	↓	1	1	$\bar{q}_n$	q_n

There is a wide variety of these kinds of flip-flops, for instance one can have flip-flops with only one asynchronous input or types where one or more inputs are internally inverted. Some ICs contain two JK flip-flops with common clock and asynchronous inputs, to save on pins. When selecting the appropriate type of flip-flop it is not only the number and nature of the inputs that must be taken into account but also the maximum power consumption and the maximum clock frequency. Flip-flops from the

TTL series operate at clock frequencies of up to 100 MHz, the power consumption is about 40 mW (dual flip-flop). The power consumption of CMOS components depends on the clock frequency which is often expressed in units µW per gate, per MHz. With modern technology (for instance in the case of the 0.35 µm IC-process), clock frequencies of more than 10 GHz can be achieved. Power dissipation and clock frequency vary greatly from type to type. For further details the user is referred to manufacturers' data books. An example of the full specifications for a digital component (i.e. a dual JK flip-flop) is given in Appendix B.2.2.

19.2 Logic circuits

This part of the present chapter on digital electronics contains some examples of circuits that are composed of logic elements. We shall successively discuss the digital multiplexer, the logic adder, counters and shift registers. The chapter ends with an illustrative example on the designing of a counter circuit for a specific application.

19.2.1 Digital multiplexer

A digital multiplexer has essentially the same function as an analog multiplexer, the difference being that the digital multiplexer has binary input and output signals. A digital multiplexer can be composed from combinatory elements (logic gates). Figure 19.15 shows a digital multiplexer circuit with 8 inputs d_0 to d_7. With the selection inputs s_0, s_1 and s_2, one of the 8 signal inputs is selected, the value of the selected d-input (0 or 1) is transferred to the output y. With the enable input e, the output can be

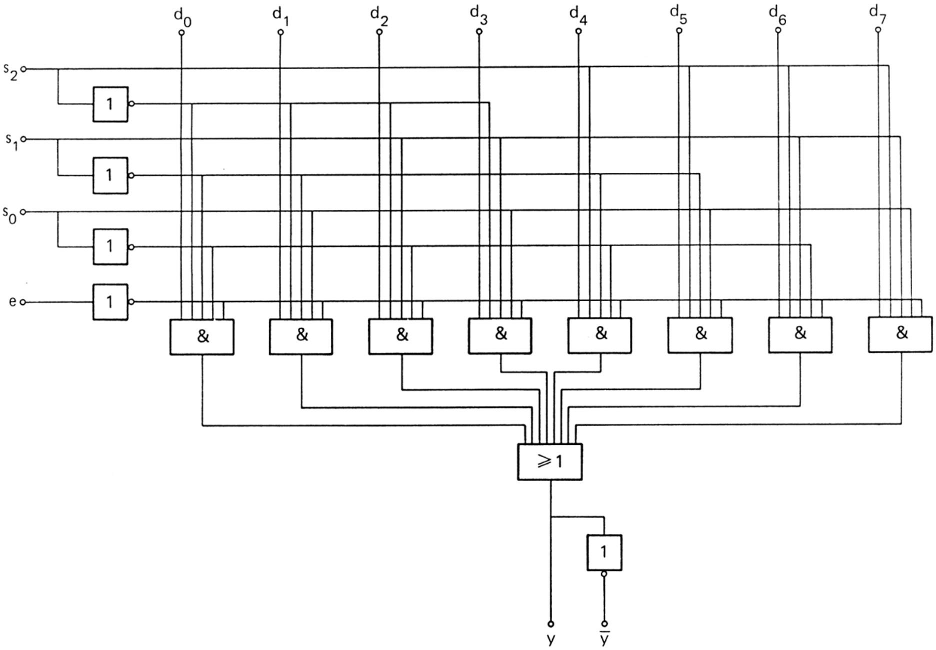


Figure 19.15. A digital multiplexer with 8 inputs d_0-d_7. The inputs s_0-s_2 are for channel selection, e is the enable input.

fixed at 0, irrespective of the selected input (compare the enable input of the analog multiplexer in Section 15.2). The truth table for this multiplexer, available as an integrated circuit, appears in Table 19.11.

Table 19.11. The truth table for the 8-channel digital multiplexer in Figure 19.15; "-" means "don't care".

e	s_2	s_1	s_0	y
1	–	–	–	0
0	0	0	0	d_0
0	0	0	1	d_1
0	0	1	0	d_2
0	0	1	1	d_3
0	1	0	0	d_4
0	1	0	1	d_5
0	1	1	0	d_6
0	1	1	1	d_7

19.2.2 The digital adder

Logic variables are either true or false and are denoted as "0" and "1" or, in short, as 0 and 1. A group of logic variables can be represented by strings of 1s and 0s, or even as binary numbers with 0 and 1 bits corresponding to the logic values 0 and 1. The logic variables then come to be considered as the value of a particular bit of a binary number. If one bears this in mind, it then becomes easy to understand the principle underlying a digital summing circuit based on binary numbers.

Figure 19.16 shows a combinatory circuit to which the two binary numbers A and B, each of which is only 1 bit, can be added. The sum is a 2-bit number since the maximum sum is $1_{10} + 1_{10} = 2_{10} = 10_2$. We may conclude from the truth table (Table 19.12) that the least significant bit s_0 of the sum is the result of EXOR operation on the variables a and b, whereas the most significant bit (here s_1) is the result of AND operation on a and b.

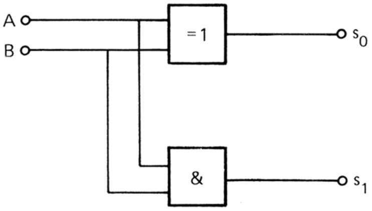


Figure 19.16. The circuit of a half-adder for the arithmetic summation of two 1-bit numbers.

The circuit seen in Figure 19.16 is called a half-adder, the variable s_0 is the sum-bit and s_1 is the carry-bit. For the summation of two binary numbers of 2 bits each, the circuit is extended to correspond to that shown in Figure 19.17. This circuit consists of a half-adder to determine the least significant bit (s_0) and the initial carry bit (c_0). Together with A and B (a_1 and b_1), which are the most significant bits, this carry bit determines the next carry bit which is, in this case, the third bit and thus the MSB of the sum. The lower part of this circuit is a full-adder. The circuit can be extended by introducing more such full-adders so that larger binary numbers can be added. The truth table (Table 19.13) shows the various binary values in the circuit for each possible combination of the two input numbers.

Table 19.12. The truth table accompanying the binary adder given in Figure 19.16.

A	B	s_1	s_0	decimal
0	0	0	0	0 + 0 = 0
0	1	0	1	0 + 1 = 1
1	0	0	1	1 + 0 = 1
1	1	1	0	1 + 1 = 2

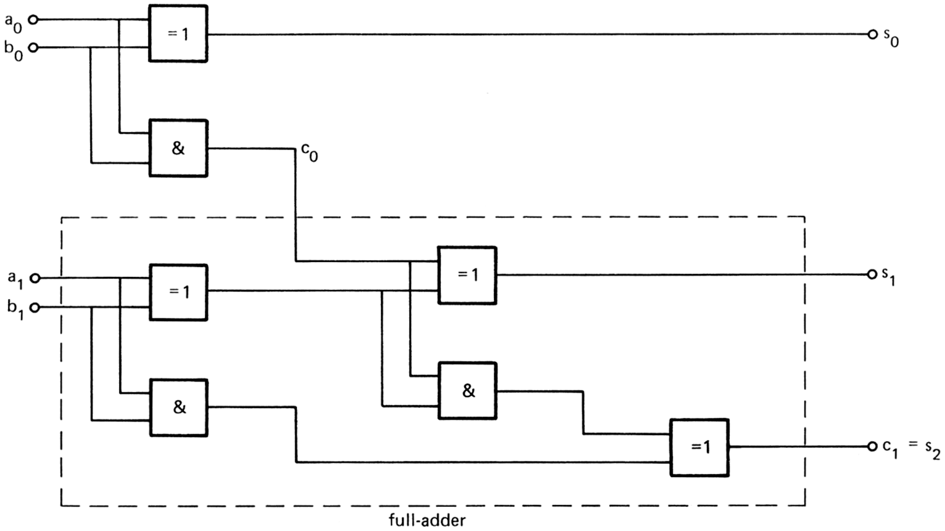


Figure 19.17. The circuit for the arithmetic addition of two 2-bit numbers. The circuit can be extended by introducing more full-adders so that larger numbers can be added.

Table 19.13. The truth table for the adder in Figure 19.17.

A		B		S			$A + B = S$ Decimal
a_1	a_0	b_1	b_0	s_2	s_1	s_0	
0	0	0	0	0	0	0	0 + 0 = 0
0	0	0	1	0	0	1	0 + 1 = 1
0	0	1	0	0	1	0	0 + 2 = 2
0	0	1	1	0	1	1	0 + 3 = 3
0	1	0	0	0	0	1	1 + 0 = 1
0	1	0	1	0	1	0	1 + 1 = 2
0	1	1	0	0	1	1	1 + 2 = 3
0	1	1	1	1	0	0	1 + 3 = 4
1	0	0	0	0	1	0	2 + 0 = 2
1	0	0	1	0	1	1	2 + 1 = 3
1	0	1	0	1	0	0	2 + 2 = 4
1	0	1	1	1	0	1	2 + 3 = 5
1	1	0	0	0	1	1	3 + 0 = 3
1	1	0	1	1	0	0	3 + 1 = 4
1	1	1	0	1	0	1	3 + 2 = 5
1	1	1	1	1	1	0	3 + 3 = 6

19.2.3 Digital counters

A counter is a digital circuit that makes it possible to count pulses, zero-crossings or periodic signal periods. Figure 19.18a shows a simple counter circuit which consists of a chain of JK flip-flops connected in series. The Q-output of each separate flip-flop is connected to the clock input of the next flip-flop. All *j* and *k*-inputs are 1, so all flip-flops act as a toggle (see Table 19.9 and Figure 19.13). Each flip-flop halves the frequency of the clock input, as demonstrated in Figure 19.18b.

The group of binary outputs corresponds to the binary coded number of clock pulses that have passed the first flip-flop. The counter is incremented one bit at each clock pulse: it is an up-counter. A down-counter is realized in a similar way but there the $\overline{q}$ outputs must be connected to the next flip-flop clock inputs, the *q*-outputs correspond to the counter output codes. Starting from the situation where all flip-flop outputs are 0, the circuit proceeds counting as follows: 0000 - 1111 - 1110 - .... - 0001 - 0000 - 1111, and so on. When devising the time diagram for such a down-counter it must be realized that the information at the *j* and *k*-inputs is stored on the clock pulse's positive edge and transferred to the output *q* at the clock's negative edge, in accordance with the master-slave principle.

Due to the time delay, each flip-flop will trigger a brief time interval later than its predecessor (this is not shown in Figure 19.18). The flip-flops do not trigger at the same time which explains why they are called asynchronous counters or ripple counters. Instead the bits change consecutively, a factor which introduces serious timing errors, in particular in the case of complex digital circuits. Synchronous counters do not have this problem. An example of a 4-bit synchronous counter is depicted in Figure 19.19. All clock inputs are connected to each other so the flip-flops trigger simultaneously. The *jk*-inputs control the flip-flops. The truth table (Table 19.14) shows for which condition $jk = 11$ (the toggle situation). The first flip-flop acts as a toggle for output a_0 while the second (output a_1) must toggle for $a_0 = 1$. So the *jk*-

inputs for this flip-flop must satisfy $j = k = q_0$. The third flip-flop (output a_2) must toggle for $a_0 = 1$ and $a_1 = 1$ which is why j and k must be equal to $q_0 \cdot q_1$. The last flip-flop (with output a_3) should toggle only if $a_2a_1a_0 = 111$, or $j = k = q_0q_1q_2$.

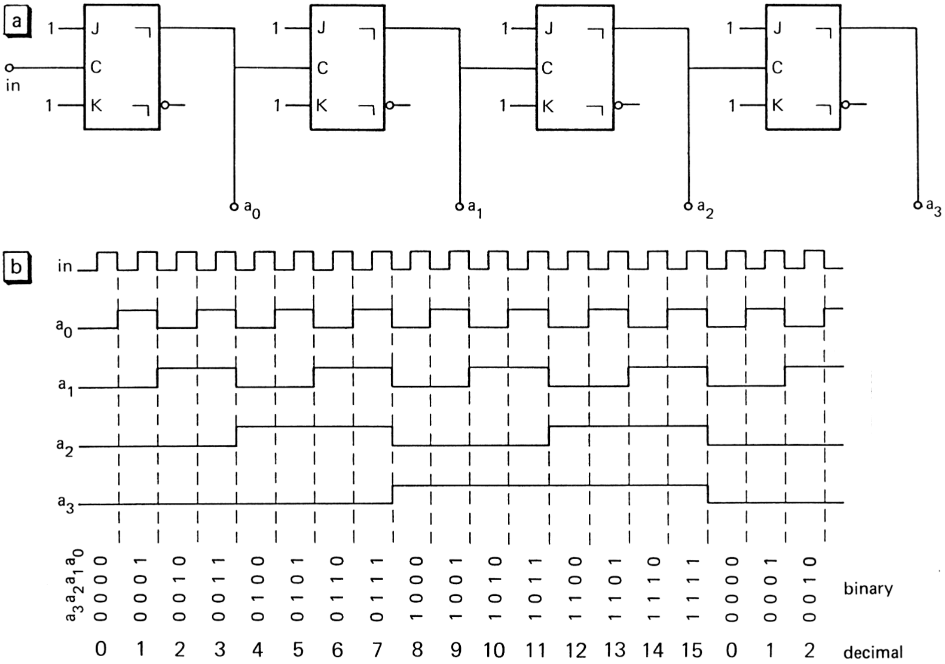

a
1
J
C
K
in
a0
a1
a2
a3
b
a3a2a1a0
0000
0001
0010
0011
0100
0101
0110
0111
1000
1001
1010
1011
1100
1101
1110
1111
0000
0001
0010
binary
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 0 1 2
decimal

Figure 19.18. (a) A four-bit binary counter composed of four master-slave JK flip-flops, each in the toggle mode, (b) a time diagram of an up-counter with the corresponding binary and decimal output codes.

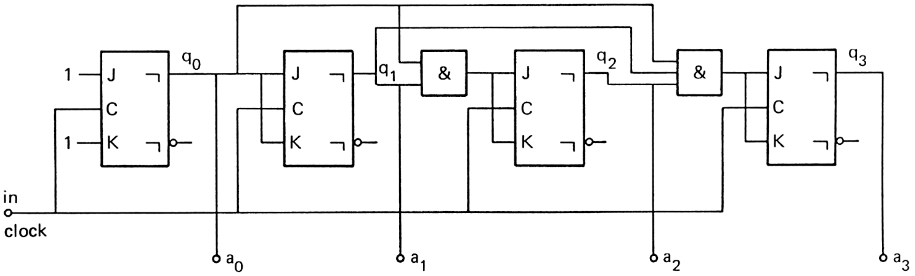

1
J
C
K
q0
q1
q2
q3
&
in
clock
a0
a1
a2
a3

Figure 19.19. A four-bit counter running synchronously with the clock.

Table 19.14. One counting cycle for the synchronous 4-bit counter of Figure 19.19.

	q_3 a_3	q_2 a_2	q_1 a_1	q_0 a_0	outputs that must change at the following clock pulse
0	0	0	0	0	
1	0	0	0	1	q_1
2	0	0	1	0	
3	0	0	1	1	q_1q_2
4	0	1	0	0	
5	0	1	0	1	q_1
6	0	1	1	0	
7	0	1	1	1	$q_1q_2q_3$
8	1	0	0	0	
9	1	0	0	1	q_1
10	1	0	1	0	
11	1	0	1	1	q_1q_2
12	1	1	0	0	
13	1	1	0	1	q_1
14	1	1	1	0	
15	1	1	1	1	$q_1q_2q_3$
0	0	0	0	0	
1	0	0	0	1	q_1

Obviously some extra gates have to be added to the string of flip-flops if synchronous operation is to be realized. Synchronous counters are available as complete integrated circuits.

There is a wide variety of counter types on the market such as: counters with a reset input (where all bits are set at zero), enable inputs (which stop the counter), counters that can be loaded with arbitrary values (preset) and counters that can count both up and down. Some 4-bit counters count from 0000 to 1001 (0 to 9) rather than from 0000 to 1111 (0 to 15). These counters, known as binary counters, usually have an extra output that indicates that the counter state is 9.

In order to be able to make the proper application choices the designer should consult the data books for details on the various types, specifications and maximum frequencies and power dissipation.

19.2.4 Shift registers

A shift register can store digital information and transfer that information when commanded by a clock pulse. Shift registers are encountered in all kinds of computers and digital signal processing equipment, for instance for arithmetic operations.

Example 19.1

The decimal number 45 (101101 in binary form) is doubled by shifting the bits one position to the left (and placing a 0 in the free position furthest to the right). The result is 1011010_2*, which is equal to* 90_{10}*. Dividing by two corresponds to shifting the bits one position to the right (to fill up the empty place with 0). The result is* $010110_2 = 22_{10}$*. The missing LSB corresponds to an arithmetic rounding-off.*

Like a counter, the shift register is composed of a whole string of flip-flops but this time the q and $\bar{q}$-outputs are connected to the j and k-inputs of the next element. The clock inputs are all connected to each other and the circuit operates synchronously (Figure 19.20).

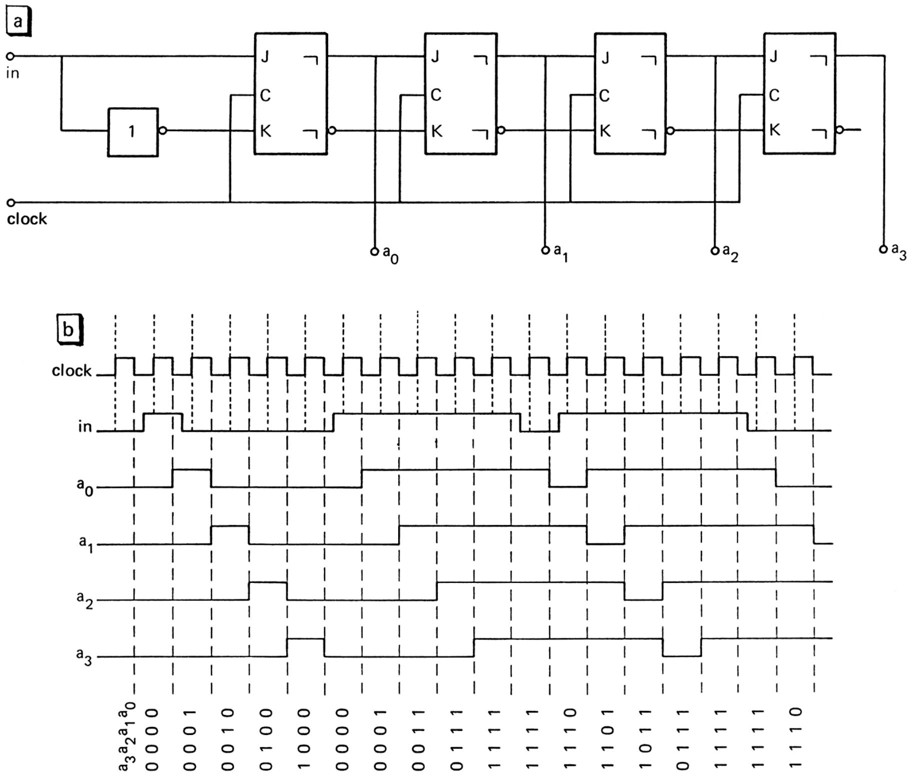


Figure 19.20. (a) A 4-bit shift register composed of four master-slave JK flip-flops, (b) a corresponding timing diagram with a shifting 1 and a shifting 0. .....=input change, - - - -=output change.

As a result of this particular coupling, each flip-flop is loaded either by a 1 ($jk = q\bar{q} = 10$) or by a 0 ($jk = q\bar{q} = 01$). If the input is 1 at the first clock pulse and 0 for the rest of the time then this 1 shifts through the register. At each clock pulse it will move one position to the right and at the end it will disappear (Figure 19.20b). The shift register of Figure 19.20 can be loaded serially so that bit after bit a binary word shifts into the register via the left flip-flop. The readout can either be done in parallel (a_0-a_3) or serially (via a_3). There are shift registers that can be parallel-loaded as well, which is faster than serial loading.

With respect to loading and reading out there are four types of shift registers that can be distinguished: serial in - serial out, serial in - parallel out, parallel in - serial out and parallel in - parallel out. Some registers can combine serial and parallel operations. Most of the available shift registers have 4 or 8 bits and some can shift to the left as well as to the right.

Apart from being used for arithmetic binary operations, shift registers are also used in communication between digital instruments, for instance between a computer and a terminal (a monitor). The link is a serial data path unlike with the computer and the terminal that operate with parallel words. If a binary word is to be transmitted it is first parallel loaded in a shift register then from there the bits are transported serially. At the receiver, the bits are loaded serially into the shift register and the word is read out in parallel. From the point of view of the instruments the communication seems to be parallel.

Figure 19.21 shows the pin connections for an integrated 4-bit bi-directional universal shift register. This kind of shift register can shift to the left as well as to the right. Other features are parallel loading, parallel readout, reset (all outputs are 0) and hold (the output remains unchanged). Table 19.15 is the function table for this shift register. This integrated circuit is available in TTL and in CMOS technology.

Table 19.15. The function table for the shift register shown in Figure 19.21. q_{ij} means: q_i at time j. "-" means 0 or 1.

	Inputs					Outputs			
Function	c	mr	s_1	s_0	d_i	$q_{0,n+1}$	$q_{1,n+1}$	$q_{2,n+1}$	$q_{3,n+1}$
Reset (clear)	–	1	–	–	–	0	0	0	0
Hold	–	0	0	0	–	$q_{0,n}$	$q_{1,n}$	$q_{2,n}$	$q_{3,n}$
Shift left	↑	0	1	0	–	$q_{1,n}$	$q_{2,n}$	$q_{3,n}$	d_{sl}
Shift right	↑	0	0	1		d_{sr}	$q_{0,n}$	$q_{1,n}$	$q_{2,n}$
Parallel load	↑	0	1	1	d_i	d_0	d_1	d_2	d_3

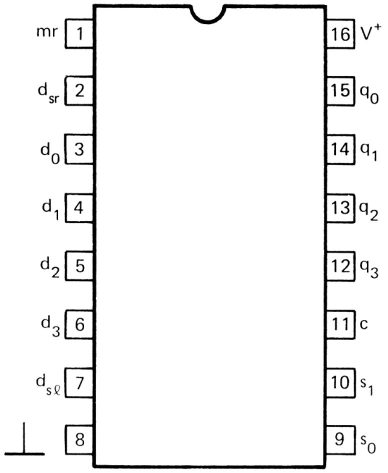


Figure 19.21. The pinning diagram for a 4-bit shift register. mr=master reset, d_{sr}=serial data input for shifting to the right, $d_0 \ldots d_3$=parallel inputs, d_{sl}=serial data input for shifting to the left, $\perp$ = ground, V^+=power supply voltage, $q_0 \ldots q_3$=parallel outputs, c=clock input, s_1,s_2=selection inputs (Table 19.15).

19.2.5 An application example

In this section we explain how digital components can be used in particular applications. For example, let us take bottles on a conveyor belt that have to be deposited in a crate. The objective is to design a digital circuit that gives off a signal when 12 bottles, corresponding to a full crate, have passed a certain point. In addition to that the number of bottles that have passed must be indicated on a display panel. What is therefore required is a counter that counts from 1 to 12 and starts again at 1. Every time a bottle passes a certain point, the counter is incremented by 1 step. A sensor (for instance a photo-detector, see Section 7.2) gives a binary signal that is 0 when there is a bottle and 1 when there is an empty place. This signal is used as a clock signal for the counter circuit.

Apparently what is needed to control the display panel is a counter, a display component and a circuit. After having consulted several data books we settled for (for instance) the digital components shown in Figures 19.22 and 19.23.

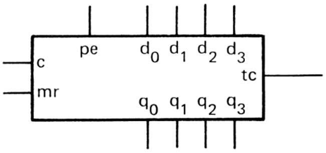


Figure 19.22. A synchronous decimal counter with possibilities for the synchronous parallel loading of a 4-bit number in accordance with Table 19.16.

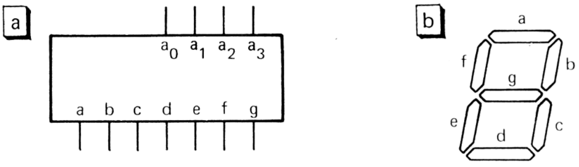


Figure 19.23. (a) a decoder for conversion from the binary code to the seven-segment code, (b) a seven-segment display unit consisting of 7 independent light sources.

Table 19.16. The functional operation of the decimal counter given in Figure 19.22; "(a)" means 1, only for a count of 9, "-" means 0 or 1, does not matter.

Function	mr	c	pe	$d_{i,n}$	$q_{i,n+1}$	tc
Reset	1	–	–	–	0	0
Parallel load	0	↑	1	0	0	0
	0	↑	1	1	1	(a)
Count	0	↑	0	–	Count	(a)

The circuit shown in Figure 19.22 is a synchronous decimal counter (Section 19.2.3) with possibilities for synchronous parallel loading. Table 19.16 shows this counter's functional operations.

The binary number to be loaded is connected to the inputs d_0-d_3. If *pe* (parallel enable) is 1 then the counter output $q_0q_1q_2q_3$ becomes equal to $d_0d_1d_2d_3$ at the next rising clock pulse. When the input is *mr* (master reset) the counter is asynchronously reset to 0000. The output *tc* (terminal count) is 1 when the counter output is 1001_2 (decimal 9), this makes it possible for a second counting section to be triggered for the tens.

The decoder for the conversion of a binary number to the seven-segment code is given in Figure 19.23a. The output signals *a* to *g* each activate one segment of the display in the way shown in Figure 19.23b.

The truth table of this decoder is given in Table 19.17. Most seven-segment decoders can be connected to an LED display directly or via resistors.

Table 19.17. The truth table for the binary to seven-segment decoder depicted in Figure 19.23a.

a_3	a_2	a_1	a_0	a	b	c	d	e	F	g	display
0	0	0	0	1	1	1	1	1	1	0	0
0	0	0	1	0	1	1	0	0	0	0	1
0	0	1	0	1	1	0	1	1	0	1	2
0	0	1	1	1	1	1	1	0	0	1	3
0	1	0	0	0	1	1	0	0	1	1	4
0	1	0	1	1	0	1	1	0	1	1	5
0	1	1	0	1	0	1	1	1	1	1	6
0	1	1	1	1	1	1	0	0	0	0	7
1	0	0	0	1	1	1	1	1	1	1	8
1	0	0	1	1	1	1	1	0	1	1	9
1	0	1	0	0	0	0	0	0	0	0	
1	0	1	1	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	
1	1	0	1	0	0	0	0	0	0	0	
1	1	1	0	0	0	0	0	0	0	0	
1	1	1	1	0	0	0	0	0	0	0	

The design that has been given so far is depicted in Figure 19.24. The counter outputs are connected to the decoder inputs that control the display LEDs. The seven LEDs have a common terminal that must be connected to ground, in series with the other terminals the resistors limit the control current to a maximum value (as a precautionary measure).

This circuit has the following counting sequence: 0, 1, 2, ..., 8, 9, 0, ... As the system must count to 12, an extra display element is required. We chose a display just for a 1, which can simply be controlled by a JK flip-flop and a buffer. The flip-flop should generate a 1 when the clock goes up during a counter output of 9. For this purpose, it is the *tc* output that is used. The flip-flop must be of the master-slave type, the information on the *tc* is stored on the negative edge of the clock pulse and appears at the output, on the positive edge of the clock. This whole operation is shown in Figure 19.25, which is the counter circuit for Figure 19.24 with an extension for counting up to 19.

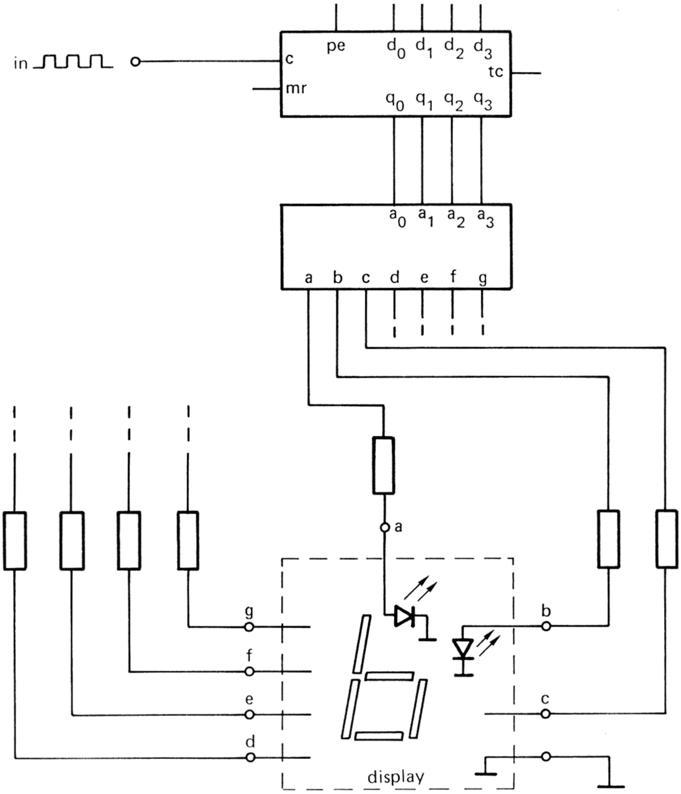


Figure 19.24. A decimal counter with a display for counting from 0 to 9.

With the timing diagram given in Figure 19.25b it appears that the circuit first counts from 0 to 19. As the flip-flop is not reset, the next cycles run from 10 to 19. The display for the tens has to be switched off after a count of 12. This is achieved by resetting the flip-flop to a count of 12 in a similar way to the count 9 setting. If we look at the time diagram we see that at a count of 12 there is a unique combination: $q_2 = 1$ and $q_A = 1$. In that case $q_2 \cdot q_A$ is a suitable signal for switching off the display unit (see the AND-gate in Figure 19.26).

With this additional circuit the counting sequence is 0,.., 11, 12, 3, 4, ... To reset the counter to 1 (after a count of 12), we utilize the parallel load feature of the IC counter. For that purpose, the inputs $d_0d_1d_2d_3$ are made equal to 1000 (d_0 is the LSB). When *pe* = 1, this number is loaded into the counter at the next clock pulse. This may only happen after having counted up to 12. In fact there is already a signal that satisfies this condition: the *k*-input for the flip-flop. Unfortunately, though, this signal cannot be used for the *pe* without introducing timing problems. For the sequential circuits to operate properly their inputs must be stable for a specified time interval before and after the clock pulse. The *k*-input changes as soon as the counter output is no longer 12, whereas the *pe* signal must remain stable when the rising clock count is changing the counter output from 12 to 1.

A signal that satisfies these conditions will be derived from an extra flip-flop triggering on the negative edge of the clock pulse (not a master-slave flip-flop). The *j*-input is connected to a signal that is 1 during the counter output of 12 whereas the *k*-input is connected to a signal that is 1 for counter outputs of 1 and 0 at 12, for instance q_A (see the timing diagram in Figure 19.26b). This *pe* signal can ultimately be used to indicate a full crate, another suitable signal for that purpose is $z = q_2q_A$ (see the timing diagram).

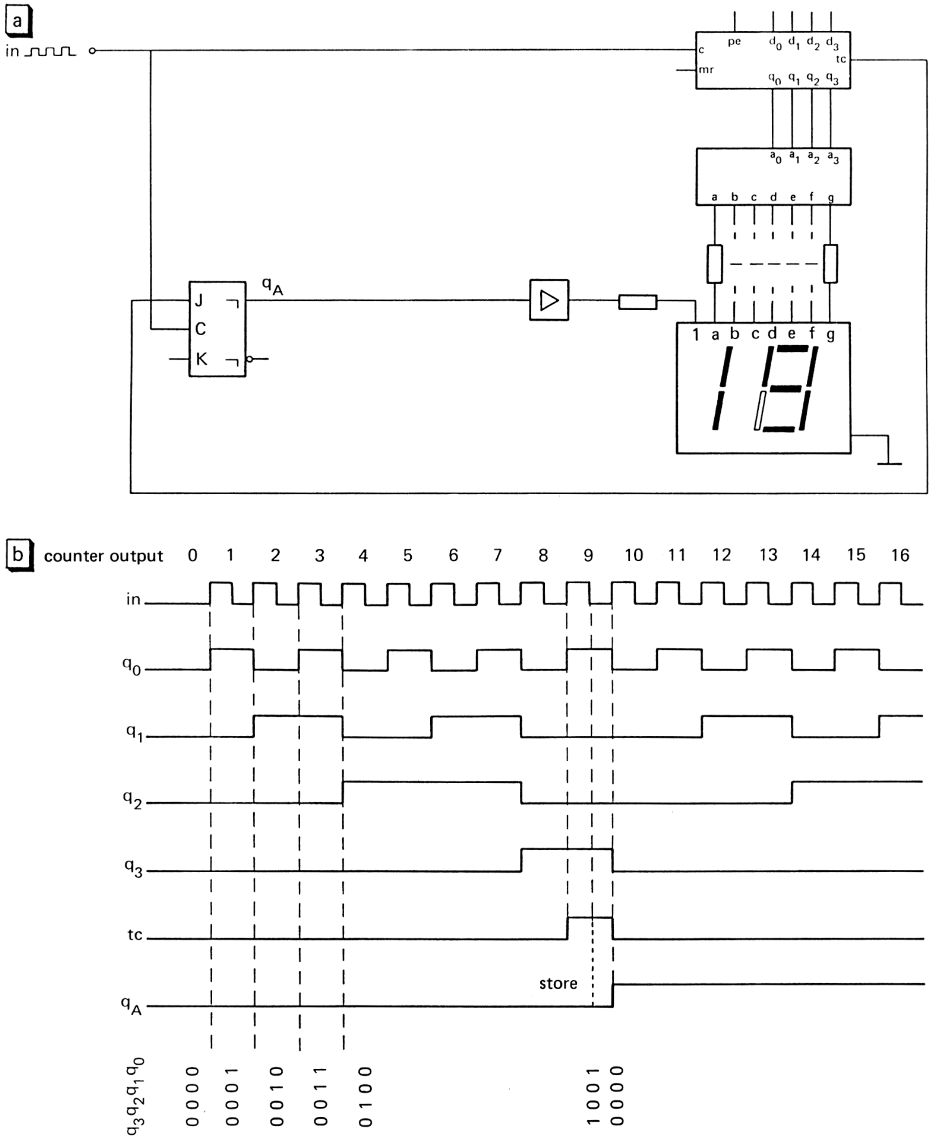

a
in
c
pe
d_0 d_1 d_2 d_3
tc
mr
q_0 q_1 q_2 q_3
a_0 a_1 a_2 a_3
a b c d e f g
q_A
J
C
K
1 a b c d e f g
b
counter output 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
in
q_0
q_1
q_2
q_3
tc
store
q_A
$q_3q_2q_1q_0$
0000
0001
0010
0011
0100
1001
0000

Figure 19.25. (a) The counter circuit shown in Figure 19.24 with the extra dimension of a display for the tens. This circuit counts to 19, (b) a corresponding time diagram showing the control for the extra flip-flop.

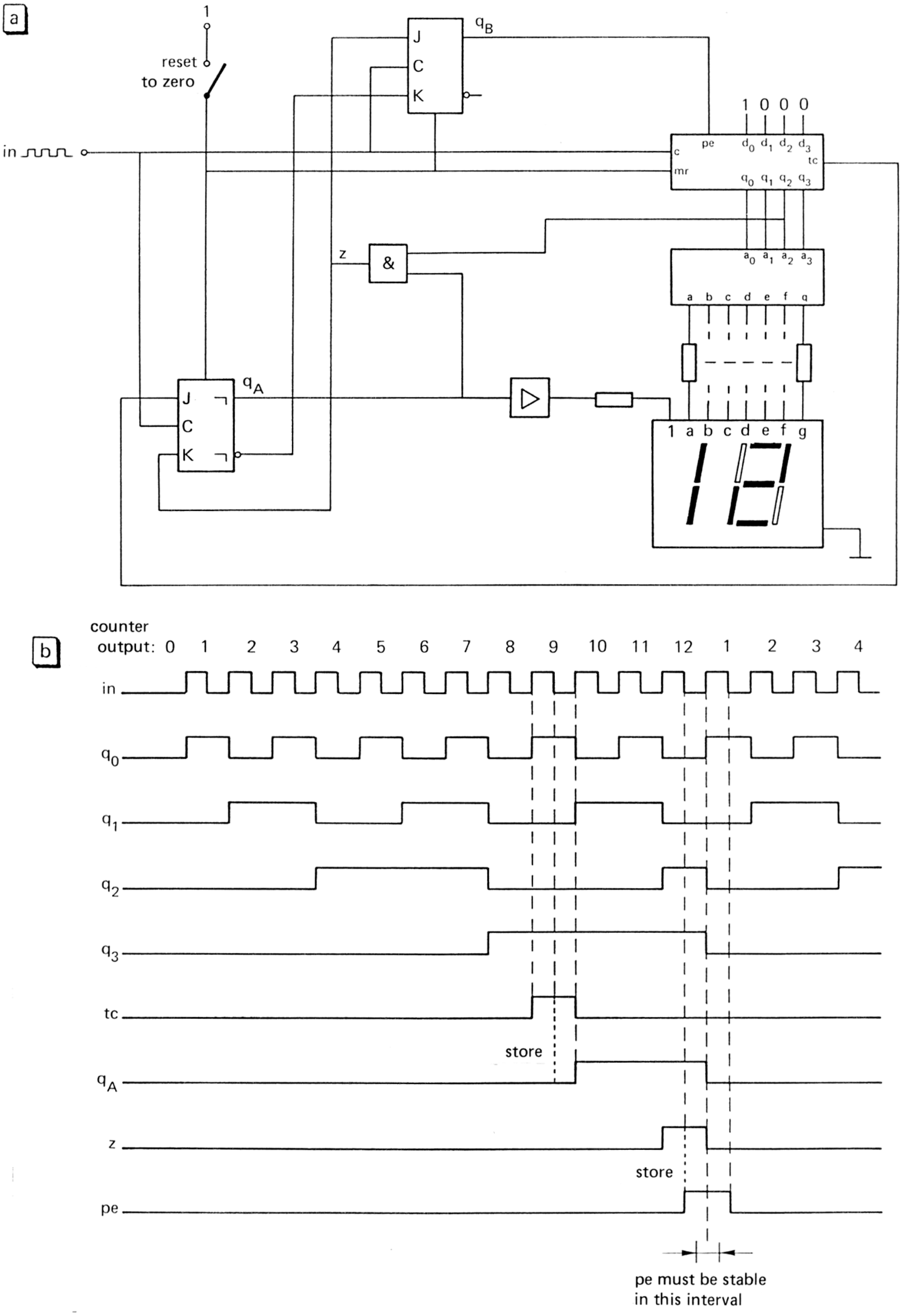

a
1
reset
to zero
in
J
C
K
q_B
1 0 0 0
c
pe
d_0 d_1 d_2 d_3
tc
mr
q_0 q_1 q_2 q_3
z
&
a_0 a_1 a_2 a_3
a b c d e f g
J
C
K
q_A
1 a b c d e f g
b
counter
output: 0 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4
in
q_0
q_1
q_2
q_3
tc
store
q_A
z
store
pe
pe must be stable
in this interval

Figure 19.26. (a) The complete digital counter circuit that counts from 1 to 12, (b) the corresponding time diagram, showing all the relevant control signals.

To reset the counter at an arbitrary moment the reset inputs of both flip-flops and the decimal counter have to be connected to a reset switch (i.e. a push button). As long as this switch is on the counter output remains zero, the resetting is an asynchronous signal.

Except for the reset aspect, the designed counter is a fully synchronous circuit. It is easy to create an asynchronous design, in fact that might even be simpler. The advantage of synchronous operation, however, is that all signal transitions occur at fixed moments, thus simplifying the design work, particularly in the case of highly complex systems.

Naturally there are various alternative approaches to design. In the first place there is programmable logic which makes use of chips with a large number of different gates and other digital components. Once the design has been completed and loaded into the computer the programmable chip is automatically configured in accordance with the design by making proper (electronic) connections between the selected components. Although in this way only a fraction of the available components may be effectively used, the time required to construct the hardware circuit is drastically reduced. In addition to this, functionality can be checked beforehand by simulating the circuit before realizing the design. Another possible approach is to use a microprocessor or microcomputer. The functionality required can be completely realized in software. This approach is preferred when the logic functions are far more complex than in the simple example of the 1-to-12 counting system given here.

SUMMARY

Digital components

- Definitions of digital signals and digital processing systems are based on Boolean logic variables algebra. A logic variable has only two values expressed as: true or false, T or F or as 1 or 0.
- Relations between two or more logic variables are represented by logic equations or truth tables.
- The four basic logic operations are AND, OR, NOT and EXOR. The digital circuits that realize such operations are the logic gates.
- The output of a combinatory digital circuit simply depends on the combination of actual inputs.
- The output of a sequential digital circuit not only depends on the actual input combination but also on earlier input values (memory function). A flip-flop is an example of such a circuit.
- An SR flip-flop has three states: set (sr = 10, output 1), reset (sr = 01, output 0) and hold (sr = 00, output remains unchanged).
- A JK flip-flop has, besides the three SR flip-flop operation modes, a fourth mode known as inversion or toggle (jk = 11). In this mode the flip-flop behaves like a frequency divider (factor 2).

Logic circuits

- Digital circuits can operate in synchronous or asynchronous ways. In the case of synchronous circuits all flip-flops trigger at the same time and at the command of a clock pulse but in asynchronous circuits this is not the case.
- Examples of circuits composed of combinatory elements (i.e. logic gates) are the digital multiplexer and the binary adder.
- A digital counter is a circuit composed of flip-flops which can count numbers of clock pulses. There are synchronous and asynchronous counters. Decimal counters are binary counters that count up to 10.
- A shift register performs division or multiplication by 2 by shifting the chain of bits to the right or to the left as required. It is also used to convert parallel words into serial words and vice versa.
- In order to control the seven-segment display, special digital circuits are available, that are called 7-segment decoders.

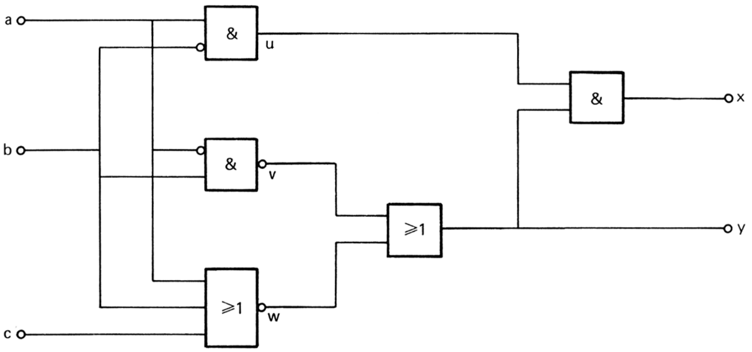
EXERCISES

Digital components

19.1 Simplify the following logic equations.
$x + xy$
$\bar{x} + xy$
$\bar{x} + \bar{x}\bar{y}$
$x(x \oplus y)$
$x(y + z) + x\bar{y}\bar{z}$

19.2 Make a truth table for the following logic relations.
a. $(a \oplus b)(a \cdot b)$
b. $(a + \bar{b} + \bar{c})(\bar{a} + \bar{b} + c)(b + c)$

19.3 On the basis of the combinatory circuit given in the Figure below, create a truth table for this circuit.



19.4 The truth table below belongs to a D flip-flop. Give a circuit diagram for such a flip-flop with just one JK flip-flop and some logic gates.

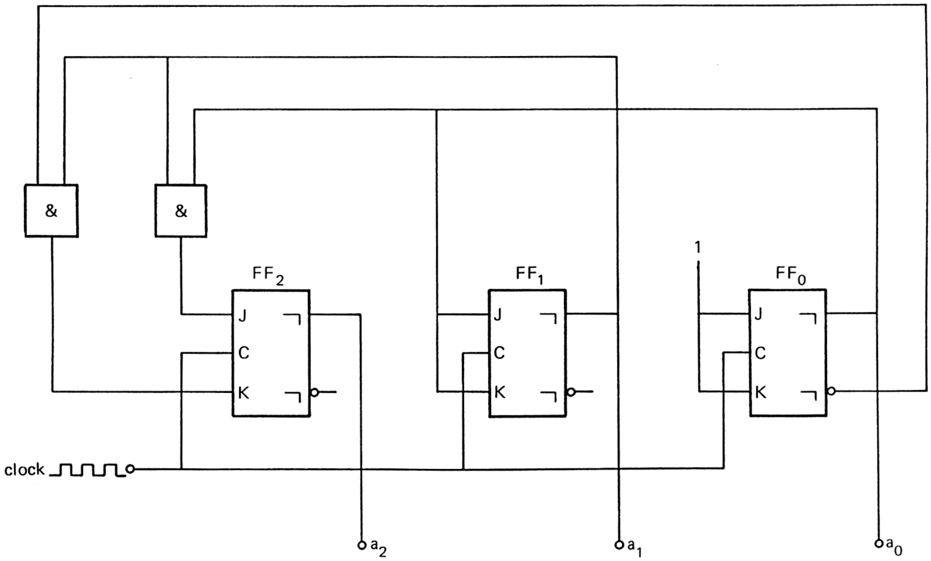
d_n	q_n	q_{n+1}
0	–	0
1	–	1

Logic circuits

19.5 Create a truth table for the multiplexer in Figure 19.15 bearing in mind that the output y is a function of the inputs d_n, the control inputs s_0, s_1 and s_2, and the enable input e.

19.6 Logic gates exhibit time delay. Discuss the effect that time delay has on large binary numbers in full adders.

19.7 The output of the counter depicted in the Figure given below is $a_2a_1a_0 = 011$. Establish what is the counting sequence.



20 Measurement instruments

Measurement systems are available for almost any electrical phenomenon. Such instruments are inevitably based on the electronic principles and circuits described in preceding chapters. In the first part of this chapter we shall review several of the more commonly used measurement instruments such as: multimeters, oscilloscopes, signal generators and analyzers whilst in the second part we shall go on to deal with computerized measurement systems.

20.1 Stand-alone measurement instruments

The way in which a measurement system is constructed will largely depend on the application for which it is built. Two main types of measurement systems can be distinguished.

When it comes to the matter of prototyping and testing it is often *stand-alone* measurement systems that are required which can be used in flexible and interactive ways. The measurements will be passed on to a person and so it will be a human-machine interface that will make up part of the system.

Any automated control system requires a measurement system that is able to perform dedicated measurement tasks. The measurement system will be *embedded* in the control system. Its output will be a signal that is used to control the system without human intervention so that no human-machine interface is required.

The classic stand-alone measurement instrument is a casing containing all the hardware, the user interface and a number of terminals needed to connect the instrument to the measurement object. The user interface will contain a menu that may be simple or sophisticated and which will allow the user to configure the instrument. The number of options available will be restricted though as it will be the manufacturer who will determine the instrument's functionality. When different quantities need to be measured in different areas then the measurement system may comprise a set of instruments (power supplies, generators, meters), all connected to the measurement object. The information required is obtained from individual instrument readings.

A more versatile approach would involve connecting all measurement devices to a PC. Such a set-up facilitates the automatic conducting of complex and large numbers of measurements. The PC in question can then be programmed, the configuration adapted and the measuring devices added or removed at will. Such systems will be examined in Section 20.2.

20.1.1 Multimeters

Multimeters are the most frequently used types of electronic instruments. They are low-cost and easy to implement. Multimeters are suitable for the measuring of voltages, currents and resistances. The input voltage is converted into a digital code and displayed on an LCD (liquid crystal display panel) or an LED display panel. Input currents are first converted into a voltage, and resistance measurements are made by feeding accurately know currents generated within the instrument into the unknown resistance before going on to measure the voltage.

In their voltage mode, electronic multimeters have a high input impedance of around 1 MΩ. Most multimeters have auto-scaling and auto-polarity features which means that the polarity and the units are displayed alongside of the measurement value. Auto-ranging and auto-polarity are realized by introducing comparators and electronic switches (reed switches or FETs).

20.1.2 Oscilloscopes

When it comes to the matter of testing electronic circuits, the oscilloscope is an indispensable instrument. It projects otherwise invisible electronic signals, notably periodic signals. Some oscilloscopes are also able to display non-periodic signals or transients.

The display section of a classical oscilloscope consists of a cathode ray tube, two pairs of deflection plates and a phosphorescent screen (Figure 20.1). The cathode, which is a heated filament, emits a beam of electrons that is directed towards the screen and produces a light spot where the electrons hit the screen. Without deflection the spot just hits the middle of the screen. The beam can be independently deflected in horizontal and vertical directions (in x and y-directions) by applying a voltage to the deflection plates. The position of the light spot changes in relation to the voltage on the plates. Whenever an amplitude-time diagram needs to be produced on the screen, the x-plates are connected to a ramp voltage generated internally in the instrument. This time-base signal sweeps the spot from left to right at a constant speed, thus resulting in a horizontal line on the phosphorescent screen. The speed can be varied stepwise to obtain a calibrated time scale.

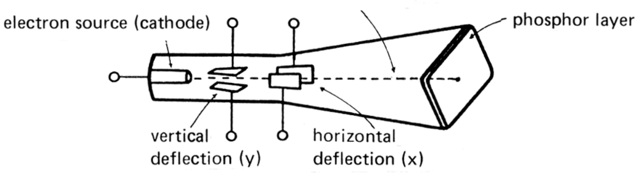


Figure 20.1. The basic structure of a cathode ray tube in an oscilloscope.

The signal to be observed is connected to the y-plate by means of an adjustable gain amplifier, thus enabling the spot to move in a vertical direction. It is the combination of horizontal and vertical displacement that results in an image of the amplitude-time diagram on the screen.

When the beam gets to the end of the scale, the time-base signal returns to the left within a very short space of time and while returning the beam is switched off (or blanked) to make it invisible to the observer. This process is periodically repeated. Due to the inertia of the human eye and the afterglow of the phosphor, the observer gets the impression that he is fully viewing part of the periodic signal.

If the frequency of the periodic input signal is an exact multiple of the time base the image on the screen will be stable (Figure 20.2). If the number of periods does not fit into the ramp interval the picture will end up passing over the screen which will hamper observation. To obtain a stable picture, irrespective of input frequency, the time-base signal will be triggered upon the command of the input. A comparator (Section 14.2.1) compares the input signal with an adjustable reference level and sets off the time base when the input passes that level. The result for a sine wave signal, and for various trigger levels and time-base frequencies, is shown in Figure 20.3.

Most oscilloscopes have an automatic trigger facility that automatically finds a correct trigger level. For composite signals the user can switch a filter to trigger only on the low or high frequency components. Other features are external triggering (from a separate signal applied to an extra terminal) and delayed trigger (where the trigger is delayed for an adjustable time interval).

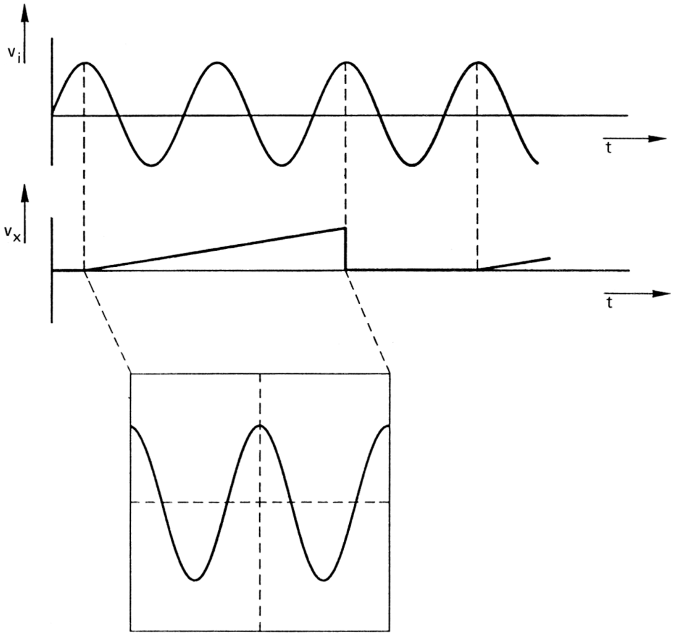


Figure 20.2. An oscilloscope image of a sinusoidal signal v_i. The picture is stable when the input frequency is just a multiple of the time-base frequency.

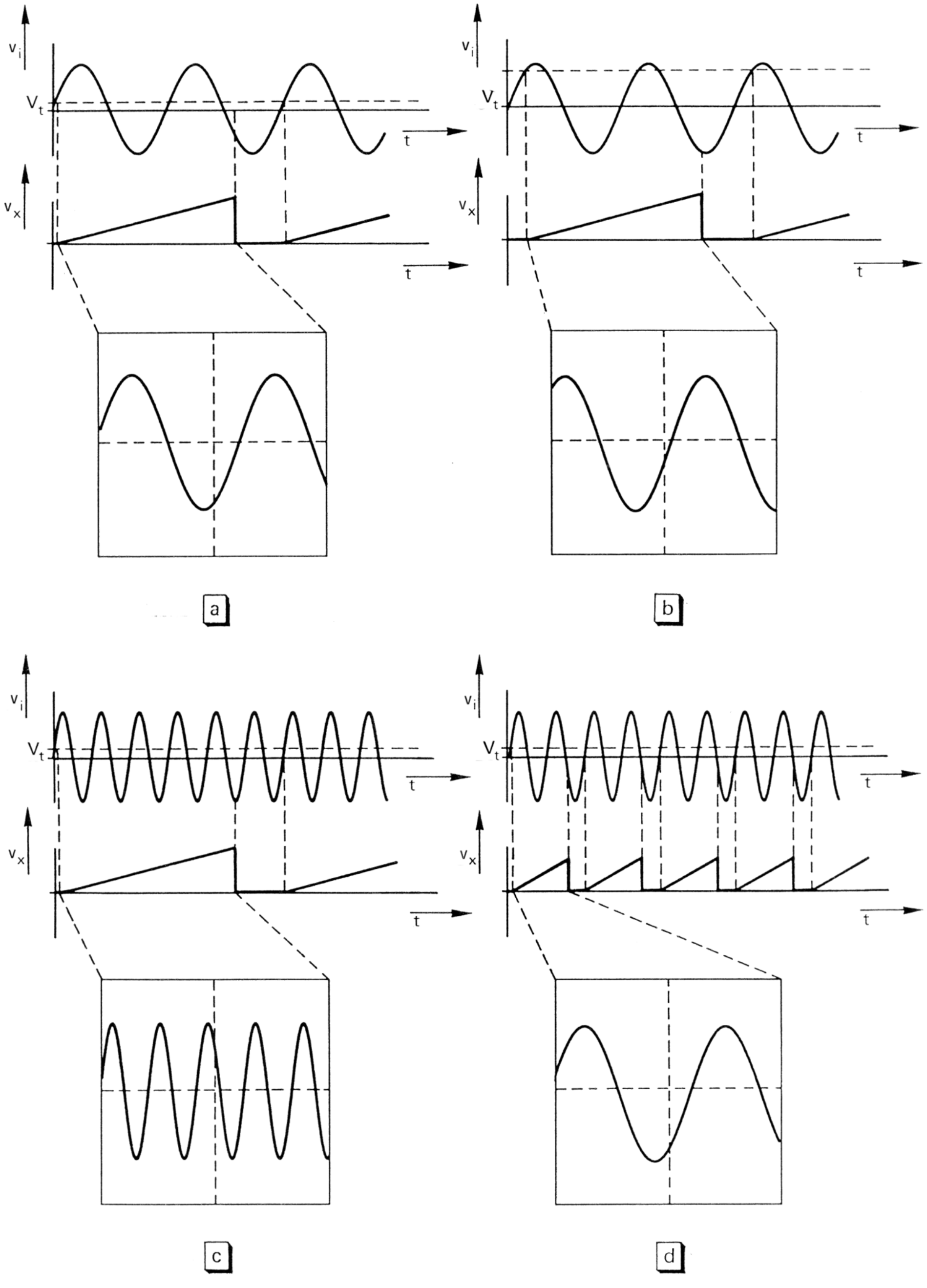


Figure 20.3. An oscilloscope image of a sine wave input signal for (a) a low trigger level v_t,(b) a high trigger level, (c) a high input frequency (with a time base and trigger level like that of (a)), (d) a high frequency with an adjusted time base frequency.

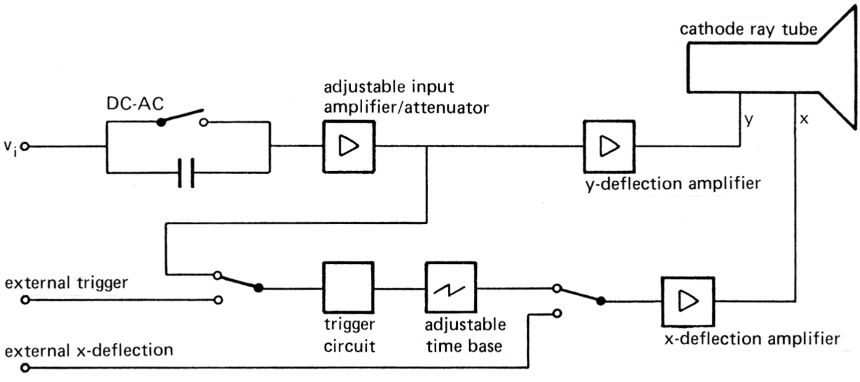


Figure 20.4. A simplified block diagram of an oscilloscope.

Figure 20.4 shows a simplified oscilloscope block diagram in which, for the sake of simplicity, the blanking circuit has been left out.

A capacitor that is in series with the input of the *y*-amplifier/attenuator thus allows the user to eliminate the DC component of the input signal. This is a useful option when, for instance, the AC component to be observed is much smaller than the average value. Most instruments have external *x*-input signal possibilities as opposed to an internal time-base. This makes it possible for, for instance, the input-output transfer (*x*-*y* characteristic) of an electronic network to be portrayed.

There are also oscilloscopes, such as dual or multi-channel oscilloscopes, which have up to 12 channels and are suitable for more than one input signal. With such instruments several signals can be observed simultaneously. There are two ways to obtain multi-channel operation with a single electron beam. The first, known as the alternating mode, involves two signals being written on the screen alternately, during a complete time-base period (Figure 20.5a). In the case of high signal frequencies, and therefore with high time-base frequencies, the images ostensibly appear on the screen simultaneously. At low frequencies the pictures clearly appear one after another and in such cases it is the chopped mode that is preferred where the electron beam switches between the two input channels at relatively high speeds (see Figure 20.5b). Each of the signals is displayed in small parts or is said to be chopped but as the chopping frequency is so high this is not visible to the observer.

Other important oscilloscope specifications are the sensitivity of the *y*-input or inputs and the maximum trace speed on the screen and therefore also the maximum input frequency. Oscilloscopes that are deployed for general use have a sensitivity of about 1 cm/mV and a bandwidth of 50 MHz. The time base of such oscilloscopes is adjustable and goes from about 50 ns/cm to roughly 0.5 s/cm. The types of oscilloscopes discussed so far operate in real time mode. Modern types (i.e. digital oscilloscopes) sample part of the input signal before presenting the stored data on the screen in a user-defined way.

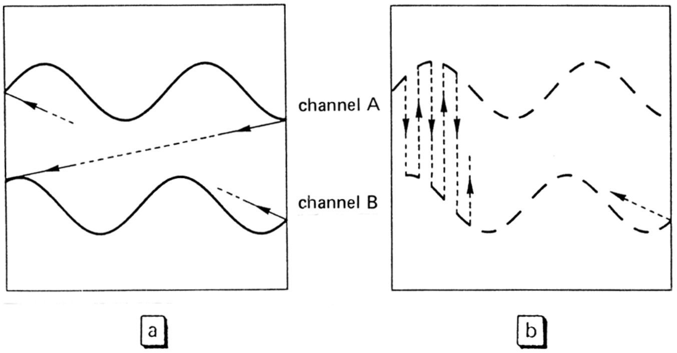


Figure 20.5. Two ways of viewing several signals with a single electron beam: (a) the alternate mode in which each input for a complete time-base period is successively shown, (b) chopped: each signal is displayed for a very short time-base period time interval.

For better measurement spot accessibility the oscilloscope is provided with one or more probes that are connected to the input channels by means of flexible cables. The capacitance of the probe cable, though, reduces the instrument's bandwidth. The input impedance of the oscilloscope is modeled using a resistance R_i and a capacitance C_i in parallel (Figure 20.6a). Furthermore, the cable capacitance is C_k and the source resistance is R_g. The voltage transfer equals:

$$\frac{V_i}{V_g} = \frac{Z_i}{Z_i + Z_g} = \frac{R_i}{R_i + R_g + j\omega R_i R_g \left(C_i + C_g\right)} \tag{20.1}$$

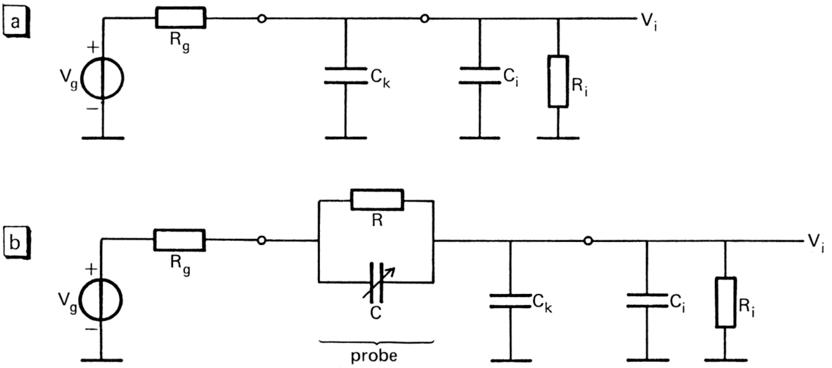


Figure 20.6. (a) A model of a voltage source and an oscilloscope with probe cable capacitance C_k and input impedance $R_i//C_i$, (b) is like (a) except that it also has a probe attenuator.

Normally, $R_g << R_i$, so the transfer approximates:

$$\frac{V_i}{V_g} = \frac{1}{1 + j\omega R_g \left(C_i + C_k\right)} \tag{20.2}$$

This clearly corresponds to a low-pass filter transfer with cut-off frequency $\omega = 1/\{R_g(C_i + C_k)\}$ (see Sections 6.1.1 and 8.1.1).

Example 20.1
Suppose that there is an input impedance of 1 MΩ // 100 pF (typical values for a standard oscilloscope), a cable capacitance of 100 pF and a source resistance of 1 kΩ. The measurement system bandwidth will then be limited to 1.3 MHz (the oscilloscope is assumed to have a much wider band). The image of a sine wave with a frequency of 650 kHz will therefore be 10% too small.

This is the reason why the probe is provided with an adjustable attenuator (voltage divider, Figure 20.6b). The signal is attenuated but the frequency transfer characteristic can be made totally independent of the frequency. The transfer for the system in Figure 20.6b is:

$$\frac{V_i}{V_g} = \frac{R_i(1 + j\omega RC)}{R_i + R + R_g + j\omega\left\{\left(R + R_g\right)\left(C_i + C_k\right)R_i + \left(R_i + R_g\right)RC\right\} - \omega^2 R_g R_i RC\left(C_i + C_k\right)} \tag{20.3}$$

In normal cases, $R_g << R$ and $R_g << R_i$, hence:

$$\begin{aligned} \frac{V_i}{V_g} &\approx \frac{R_i(1 + j\omega RC)}{R_i + R + j\omega R_i R\left(C + C_i + C_k\right) - \omega^2 R_g R_i RC\left(C_i + C_k\right)} = \\ &= \frac{R_i}{R_i + R} \cdot \frac{1 + j\omega RC}{1 + j\omega \frac{R_i R}{R_i + R}\left(C + C_i + C_k\right) - \omega^2 \frac{R_i R}{R_i + R} R_g C\left(C_i + C_k\right)} \end{aligned} \tag{20.4}$$

If one ignores the ω^2 term, which is usually allowed, the transfer will have the real value $R_i/(R_i + R)$ (so it is independent of the frequency) provided that

$$RC = \frac{R_i R}{R_i + R}\left(C + C_k + C_i\right) \tag{20.5}$$

or:

$$RC = R_i\left(C_k + C_i\right) \tag{20.6}$$

Example 20.2
A typical value for the probe attenuation is a factor of 10. With input resistance R_i = 1 MΩ (as seen in Example 20.1), the series resistance R in the probe must be equal to 9 MΩ. The corresponding value of C *is about 13 pF. The input impedance of the total measurement system (i.e. the oscilloscope with the probe) then becomes 10 MΩ // 12 pF.*

The probe attenuator allows voltages emanating from sources with a fairly high input impedance to be measured without this affecting the system's frequency response. Prior to measuring, the user must adjust probe capacitance C to frequency independent transfer. The adjustment is effected with a calibrated square-wave test signal available from an extra terminal at the front of the instrument.

20.1.3 Signal generators

When it comes to the matter of testing and tracing errors, signal generators are indispensable instruments and they come in a wide variety of types, ranging from simple sine wave oscillators (Section 16.1) to complex, processor-controlled systems that can be used for various signal shapes and signal parameters.

DC signal sources produce an accurately adjustable DC voltage or current, derived from a Zener reference source (Section 9.1). The DC voltage or current is either used as a test signal or as a stabilized power supply voltage.

Sine wave generators produce sinusoidal voltage which tends to be used for test signals, for instance to measure the frequency response of an analog signal processing system. These generators are available for frequencies ranging from very low values (down to 10^{-4} Hz) to several GHz. The basic principles of these generators are discussed in Chapter 16.

A pulse generator produces periodical square wave and pulse wave voltages with adjustable frequencies, pulse heights and pulse widths or duty cycles (see Section 16.2.1.). There are also special types of generators which produce pulses with very short rise and fall times that go down to 0.1 ns, which are used to investigate wideband systems.

Periodical signals, such as ramp, triangle and square wave signals are generated by a function generator. Some of these instruments have multiple outputs, such as triangular and square wave signals, which have equal frequencies. They are versatile instruments that are widely used in laboratory situations.

The frequency stability of the signals is determined by the stability of circuit components such as resistors and capacitors (Chapter 16). If much higher stability is required then a crystal oscillator is used. Piezoelectric elements (Section 7.2.5) are forced to resonate mechanically when AC signals are applied. The resonance frequency is almost exclusively determined by the dimensions of the crystal which is why the output frequency is as stable as the crystal. Any other frequencies that are obtained are derived from the resonance frequency by means of frequency division and by using digital circuits (see Section 19.1).

Some generators have what is known as voltage-controllable output frequency which is why they are called voltage controlled oscillators (VCOs, Section 16.2.4). In combination with second generators, the frequency can be swept linearly or logarithmically over predefined ranges (Figure 20.7). Such sweep generators are used in instruments to analyze frequency-dependent transfers and impedances.

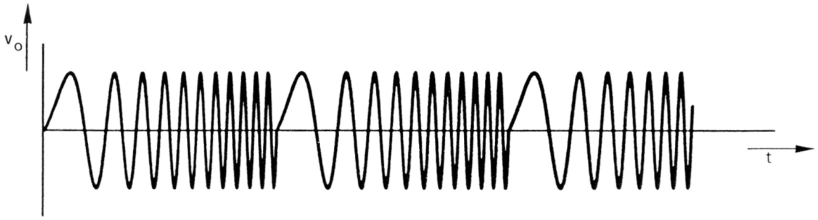


Figure 20.7. An example of a sweep oscillator output signal.

Finally we come to the matter of noise generators. These instruments generate stochastic signals that are usually derived from noise-producing elements, like for example resistors that produce white noise (Sections 2.1.3 and 5.2.2). The frequency spectrum or the noise band width is varied by using electronic filters. Noise generators that are based on digital principles generate quasi-stochastic signals that is to say, binary signals that are periodical but during any one period the transitions will be random. Noise signals are often used as test signals, the advantage over sinusoidal test signals being that noise contains a continuous range of frequency components.

All voltage generators should have low output impedance if they are to match properly (Section 5.1.3). A standard value is 50 Ω which is a value that is employed in high frequency systems as well (see characteristic impedance, Section 5.1.3). The output amplitude is continuously adjustable and ranges from 0 to 10 V or more. Some generators have adjustable DC levels (i.e. offset).

In most processor-based generators the wave form, frequency, amplitude and offset are menu-driven and determined by push buttons. In general, these instruments are far more expensive but they are more accurate and stable and certainly more flexible when compared to analog instruments.

20.1.4 Counters, frequency meters and time meters

All instruments that are used to measure numbers of pulses, frequency and time intervals are based on the counting of signal transitions. The input signal is applied to a comparator or Schmitt trigger (Sections 14.2.1 and 14.2.2) which converts the analog signal into a binary form while preserving the frequency. Such binary signals are applied directly to the digital counter circuit (Figure 20.8) which can be reset by an externally controllable reset switch. In this way, the number of input pulses can be counted. The same circuit can be used to measure input frequency, but in such cases the comparator output is only linked to the counter at specific time intervals. The counter counts the number of pulses or zero crossings per unit of time, thus also recording the frequency. The result is directly displayed in Hz. The required reference time is derived from a stable crystal oscillator; the frequency counter range not only depends on the counter range but also on the time interval for which the pulses are counted. The latter can be varied step-wise by using an adjustable frequency divider. If the range is changed in steps of 10s, the display can simply be changed accordingly by shifting the decimal point.

If this principle were also applied to low frequencies the measurement time would be too long. In such cases it is therefore more appropriate to measure the period time rather than the frequency. This can simply be done by interchanging the position of the comparator and the reference oscillator shown in Figure 20.8. The counter counts the number of pulses from the reference oscillator during half the input signal period. In much the same way the range can be changed by using the frequency divider.

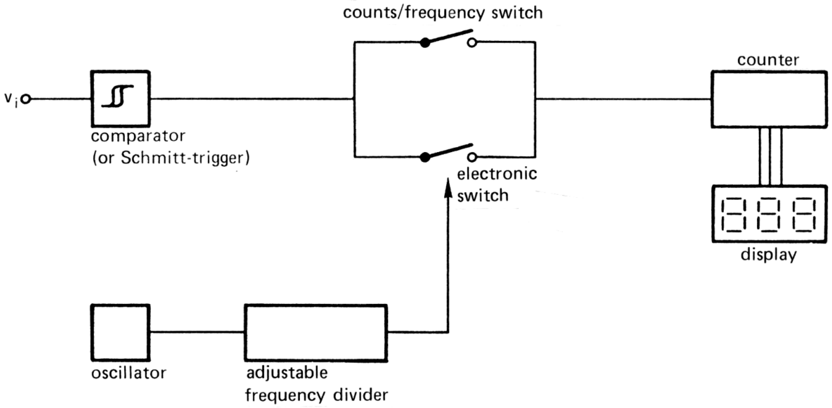


Figure 20.8. The electronic counter principle. When used as a frequency measuring instrument the pulses are counted within a defined time interval. The frequency divider is used to make adjustments.

The measurement range of general purpose frequency meters can be as high as 100 MHz, special types can even count up to several GHz. The input impedance has a high value, generally of 1 MΩ for the common types and 50 Ω for high frequency instruments.

The input signal should always exceed a specified minimum value otherwise correct conversion to a binary signal cannot be guaranteed.

20.1.5 Spectrum analyzers

A spectrum analyzer is a measurement instrument that is used to depict the frequency spectrum of a signal. Figure 20.9 gives an idea of the instrument's basic structure. The voltage-controlled band-pass filter (VCF) has a very narrow band. The position of the filter in the frequency band can be electronically varied by means of a control voltage generated by a ramp generator or simply by means of digital signal generation. During the filter sweep, the amplitude of the output signal is continuously determined by synchronous detection (Section 17.1.3). The synchronous detector behaves like a narrow band-pass filter with a central frequency that is equal to the frequency of the reference signal (Section 17.2). Information both on amplitude and on frequency is stored and displayed as a spectrum on a monitor (Figure 20.10).

20.1.6 Network analyzers

A network analyzer measures and displays the frequency characteristic or Bode plot of a two-port network. The simplified structure of a network analyzer is portrayed in Figure 20.11.

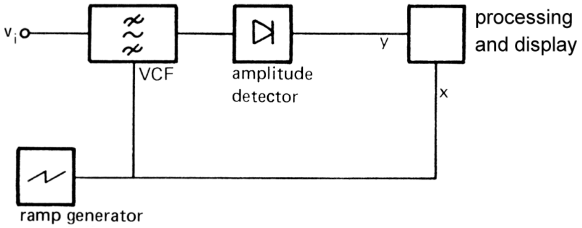

v_i
VCF
amplitude detector
y
x
processing and display
ramp generator

Figure 20.9. The basic structure of a spectrum analyzer.

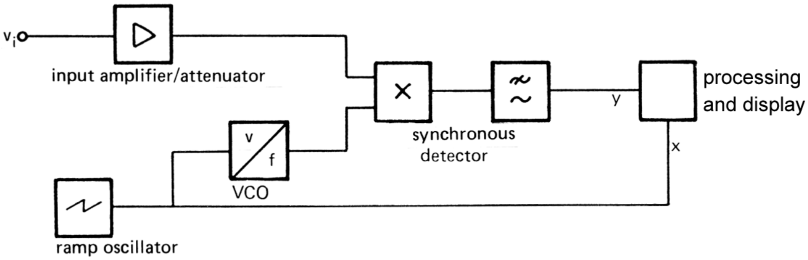

v_i
input amplifier/attenuator
v
f
VCO
ramp oscillator
synchronous detector
y
x
processing and display

Figure 20.10. The basic structure of a spectrum analyzer with a synchronous detector as a band-pass filter and an amplitude detector.

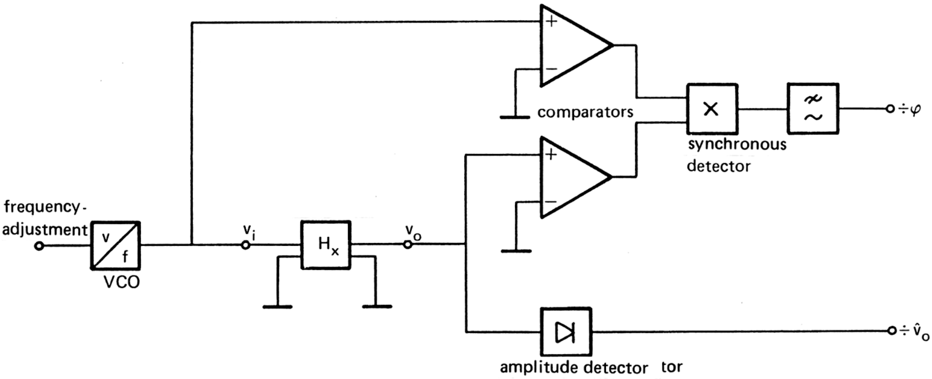

comparators
synchronous detector
$\div\varphi$
frequency-adjustment
v
f
VCO
v_i
H_x
v_o
amplitude detector tor
$\div\hat{v}_o$

Figure 20.11. The simplified structure of a network analyzer used to measure and display both the amplitude transfer and the phase transfer of a two-port network.

The VCO generates a sinusoidal voltage with controllable frequency. In manual mode the frequency can be adjusted by the user, in sweep mode the instrument varies the frequency by adding a ramp voltage to the VCO control input. When the VCO output is constant or calibrated, the amplitude detector's output becomes a gauge of the network's amplitude transfer. The phase transfer is determined by synchronous detection. A synchronous detector responds to the cosine of the phase difference and the respective amplitudes (Section 17.1). A linear phase response is achieved by converting both the input and the output signals to binary signals (square wave signals) with fixed amplitudes. The output after synchronous detection is linearly proportional to the phase difference between the input and output of the system being tested and independent of the signal amplitude.

By controlling the frequency of the VCO over a certain frequency range, the Bode plot (and ultimately the polar plot, Section 6.2) can be projected on the screen.

20.1.7 Impedance analyzers

An impedance analyzer measures the complex impedance of a two-pole network at discrete or perpetually varying frequencies. Impedance analyzers usually have a digital display or a cathode ray tube to give a complete picture. In Figure 20.12 we see the simplified structure of an instrument used to display impedance on a numerical display screen.

The VCO voltage is applied to the impedance being tested. The current through the impedance is converted into a voltage (Section 12.1.1). The amplitude and phase of this voltage is measured in a similar fashion to the network analyzer discussed in the preceding section. The instrument is equipped with AD-converters and a microprocessor to calculate, on the basis of amplitude and phase, other impedance quantities such as real and imaginary parts (Section 4.1.1) depending on how the network is modeled. The user is free to choose the model that is most suitable for the circuit in question.

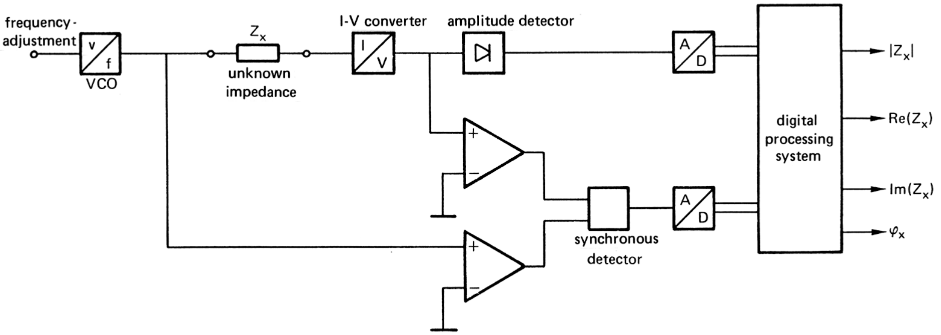


Figure 20.12. The simplified structure of an impedance analyzer with a (micro-)processor for the calculating of various impedance parameters.

20.2 Computer-based measurement instruments

Measurement systems that can be computer controlled do not differ essentially from other measurement instruments. They have special interfaces for computer communication purposes. In such cases the task of the computer is to control the instrument and the further processing of the measurement data. As the processing capacity of computers is great, several instruments can be simultaneously connected to any one computer (Section 1.1).

The PC monitor is then used for the numerical or graphical presentation of the relevant measurement data. Control buttons and instrument display can also be shown on the monitor. In that way the user is able to configure a measurement system, including various instruments, with the use of the computer and read the data as though reading from the measurement instrument itself. This practice has led to the coining of the phrase "virtual instrument" since the PC presents itself to the user as a measurement instrument. This matter will be further discussed in Section 20.2.4.

20.2.1 Bus structures

Just like in computers and processors, bus structures are used to reduce the amount of wiring and the number of interface circuits (Figure 20.13). The instruments are all connected in parallel to the bus. To transport the data in an orderly fashion, each instrument is equipped with a more or less intelligent interface.

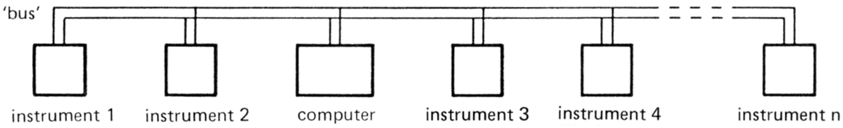


Figure 20.13. An example of a bus system for the connection of several measurement instruments to a single computer system.

A bus protocol governs the communication between the instrument and the PC and allows the instrument to be controlled and measurement data to be read from it. As is usual with bus-systems, more than one instrument can be connected to the PC across one bus. Each instrument has its own address on the bus.

One characteristic of a bus protocol is that more than one instrument may be connected to the same bus. Adding an instrument to a measurement arrangement does not therefore require that hardware additions be made to the computer. Data placed on the bus by one instrument will reach all the other instruments. However, it will only be the instrument addressed on the address bus that will read the data from the data bus and react to it. Most bus protocols are layered, for example in line with the OSI/ISO communications reference model which also forms the basis to the TCP/IP Internet protocol.

Several standardization organizations have tried to define and introduce a standard bus structure suitable for measurement instruments produced by various manufacturers. Two organizations have succeeded in establishing acceptable standard for instruments, the IEC (International Electrotechnical Commission) and the IEEE (Institute of Electrical and Electronics Engineering). The two standards adopted are the IEC 625 and the IEEE 488 bus. As these bus systems are widely used in instrumentation, we will consider them briefly in this book.

20.2.1.1 IEEE 488 and IEEE 1394

The first bus protocol to be developed specifically for instrumentation purposes was the IEEE 488. Originally defined by Hewlett & Packard in 1965 it was finally accepted as an IEEE standard in 1975. It is therefore quite an old protocol. It is also known as the HP-IB (Hewlett & Packard Interface Bus) and as the GPIB (General-Purpose Interface Bus). It contains eight parallel data lines and eight control lines. Rates up to 1 Mb/sec can be achieved. The maximum bus length is 20 m. The interface with a computer is created by means of an ISA or PCI board. Although its specifications are perfectly adequate for many instrumentation tasks, there is a tendency to demand higher data rates and lower costs while making use of a more modern protocol.

The IEEE 1394 protocol derived from "Firewire" (Apple Computer Inc.) and was defined in 1995. Often the names Firewire and IEEE 1394 are used to refer to the same protocol. Basically it is designed to enable high-speed communication between computers and "data intensive" peripherals, like cameras and scanners. It was originally intended for the consumer market. Data-rates of up to 400 Mb/sec can be achieved but the 1394a and 1394b protocols provide even greater capacities. As it is, Firewire has the drawback of not providing error correction as such vast amounts of data are often fault tolerant. "Hubs" split a Firewire connection in two, allowing flexible star and tree configurations to be formed.

20.2.1.2 The IEC 625 instrumentation bus

The internal structure of an instrumentation bus will be explained in more detail on the basis of the IEC 625 protocol. The IEC bus is restricted to a maximum of 15 measurement instruments (or, only in particular circumstances, 31). These can be instruments of various kinds, such as multimeters, frequency counters, analyzers, plotters and signal generators.

The two instrument functions that are distinguished are: "listen" and "talk". An instrument that operates as a listener can only receive messages from the bus whilst a talker can only transmit messages. Most instruments, however, are designed both for transmitting and receiving messages, for instance in order to control the measurement range and in order to transmit data. No instrument can act as a listener and as a talker at the same time. Several instruments may function simultaneously as listeners but only one instrument can talk at any given time.

To regulate the information flow along the bus one of the connected instruments, usually a computer, acts as a controller or supervisor. The controller determines which instrument is the talker and which is the receiver. Each instrument has a unique address (equipped with a series of small switches usually located on the back plane of the instrument). Only the instrument with the address that corresponds with the address transmitted by the computer receives the message. It is even possible to let two instruments communicate with each other without the intervention of the computer.

The IEC 625 bus consists of 16 lines, 8 of which are reserved for the parallel transmission of binary data while the other 8 lines, which will be described later, are there for control purposes.

The data transport is bit parallel and word serial and is transmitted byte by byte. The data consist of measurement data or commands for the selected instruments. One message may be composed of several bytes.

To guarantee proper data transport, irrespective of the differences in instrument response times, the bus contains three special control lines known as the handshake lines. These lines have the following functions.

- NRFD (Not Ready For Data): all instruments indicate by means of this line whether they are ready to accept data. The talker must wait until all listeners are ready as will be indicated by a false NRFD. Logically speaking this signal is the OR operation on all connected NRFD lines: NRFD = NRFD(1) + NRFD(2) + ...

- DAV (DAta Valid): this signal is made true by the instrument assigned to the task of talker just when all the active listeners are ready to receive data (NRFD is false). The receiving instruments know that the data on the data lines contain relevant information.
- NDAC (Not Data ACcepted): the listeners make this signal true as soon as data transmission is to take place (which is only possible in the case of DAV true). NDAC remains true as long as the last instrument has received the data from the data lines. The talker must keep the data on the data lines at least until that moment. A possible next byte can only be put on the data lines when NDAC is false. The NDAC is a logic OR function of the individual NDAC of the instruments: NDAC = NDAC(1) + NDAC(2) + ...

Figure 20.14 shows the timing diagram for the handshake procedure followed by each byte that is transmitted.

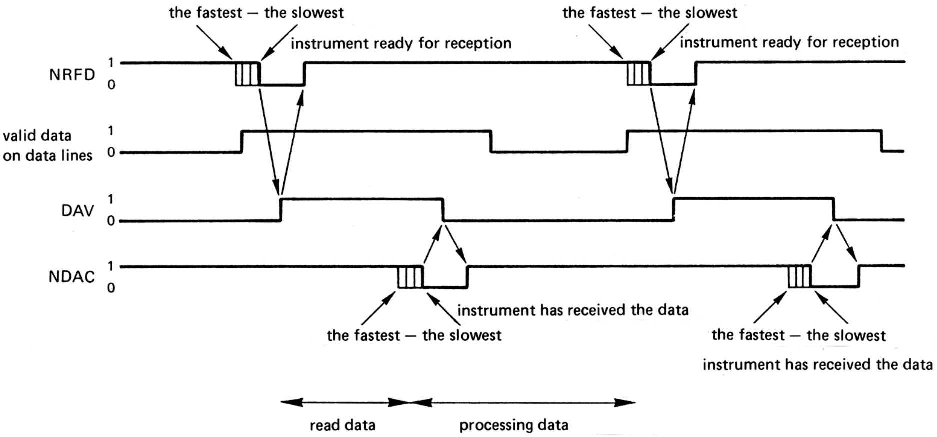

the fastest — the slowest
instrument ready for reception
the fastest — the slowest
instrument ready for reception
NRFD
1
0
valid data on data lines
1
0
DAV
1
0
NDAC
1
0
instrument has received the data
the fastest — the slowest
the fastest — the slowest
instrument has received the data
read data
processing data

Figure 20.14. The handshake procedure time diagram in the IEC-bus.

Data transport over the bus is asynchronous. Due to the handshake procedure, instruments that have diverse processing times can be connected to the bus so that the transport speed is determined by the slowest of the connected instruments.

Some bus lines are connected in what is called a wired-OR manner (Figure 20.15). The bus lines (in this case for the NRFD line) are all connected to the positive power supply voltage via a resistor. Each instrument contains a switch that can connect the lines to ground. The line voltage is only high if none of the instruments has short-circuited the line, the line is low (0) if at least one instrument has put the switch on. In this way the OR function between the NRFD lines is realized without separate OR gates, which would involve complex wiring.

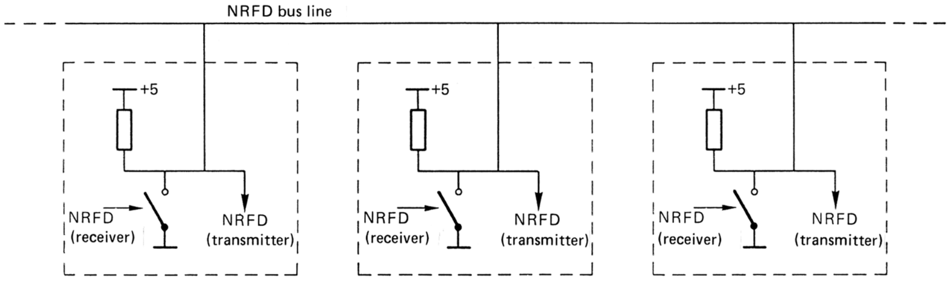


Figure 20.15. Wired-OR lines for the NRFD line. When the listener is not ready to accept data, the switch is off, the line voltage is low (0) and the talker is waiting.

The 5 remaining control lines are the management lines. They have the following functions.

- IFC (InterFace Clear): the controller can reset the bus, all connected instruments will be set in an initial state awaiting further commands.
- ATN (ATtentioN): the controller can switch from the command mode to the data mode. In the data mode, the data lines are used to transport measurement data, in the command mode instrument control signals are transported over the data lines.
- REN (Remote ENable): the controller takes over the control of the instruments, the front plate control elements are switched off.
- SRQ (ServiceReQuest): one or more instruments can signal to the controller to execute a particular program.
- EOI (End Or Identify: in the data mode (ATN=false) the EOI means that the current byte is the last byte of the message. In the command mode EOI indicates that a service request has been executed.

Normally the bus system user has nothing to do with the bus interfaces. Special ICs take care of the communication, including the handshake procedure. Instruments that are provided with an IEC bus connection can be mutually connected by means of a special cable (via what are known as piggy-back connectors). Evidently, the user has to write a program for the controller to execute the desired measurement sequence.

20.2.2 An example of a computer-based measurement system

Just to illustrate the versatility of computers in instrumentation we shall consider in this section how computer measurement systems can be applied in specific ways.

The example taken here relates to the automatic testing of a certain type of humidity sensor. The two aspects we shall examine are these: automated measuring during sensor production and automatic measuring in the development phase.

We shall take as our example the Al_2O_3 humidity sensor which consists of an aluminum substrate where the top is anodically oxidized to thus form a porous layer (Figure 20.16).

A very thin gold layer is then deposited above the oxide that is permeable to water vapor molecules. Water vapor from the air (or gas) enters the pores at a rate that is proportional to the humidity content. The sensor acts like a capacitor where the capacitance changes in accordance with the amount of water absorbed or, in relation to the relative humidity.

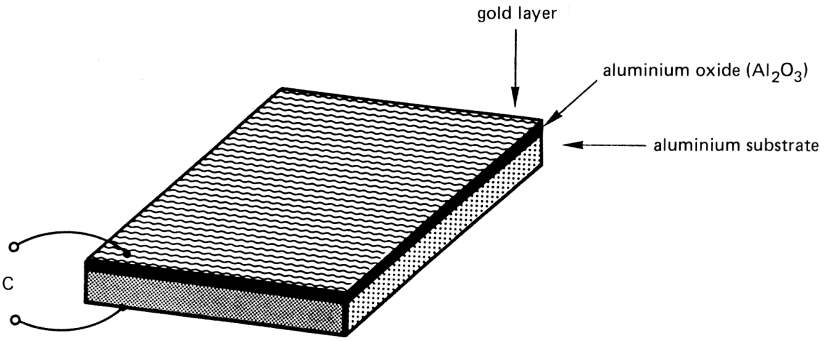


Figure 20.16. A capacitive humidity sensor. The top layer of gold is very thin and permeable to water molecules. The capacitance changes according to the amount of water absorbed.

The formation of the porous layer is a random process which means that all sensors have different absorption characteristics, even when fabricated under identical environmental conditions. The production parameters can also vary (the anodization process is easily affected by the temperature and composition of the electrolyte). All these differences result in different absorption characteristics in the sensors produced. Here one may therefore speak of capacitance as a function of humidity. This means that all sensors need to be separately calibrated if the manufacturer wants to provide precise specifications. In such cases sensors will be delivered with their own unique calibration charts or will have PROMs with calibration tables. The actual calibrating can be done automatically (Figure 20.17).

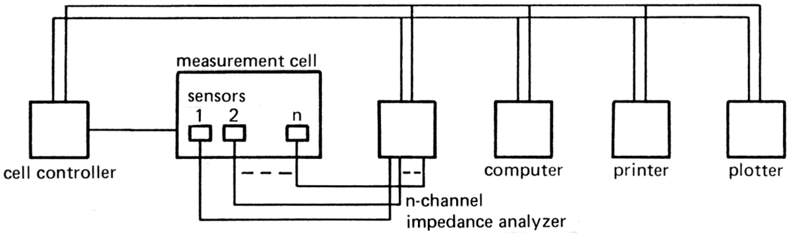


Figure 20.17. An example of a measurement arrangement for the automatic calibration of humidity sensors. The calibration curves and other specifications can be plotted automatically on the paper.

In the interests of efficiency, many sensors will be bundled together in one measurement cell (or climate chamber) where the temperature and humidity are controllable. A multi-channel impedance analyzer (Section 20.1.8) measures the capacitance of each sensor and the measurement results are stored in the computer or printed out.

Each measurement sequence starts with the initialization of the measurement instruments: the climate chamber is adjusted to the proper temperature and humidity, and the impedance analyzer to the desired frequency. The computer-controlled multiplexer in the analyzer then scans the connected sensors and the corresponding measurement data is stored in computer memory. Finally, the computer adjusts the climate chamber to another temperature or humidity. After a while (and this can take a time because of the rather long response time of the climate chamber and the sensors) the sequence is repeated. In this way, the whole humidity range can be covered. Afterwards the measurement results are printed out or a complete characteristic is plotted for each individual sensor. The chart created is provided with additional data such as the measurement frequency, the date and time of calibration and the sensor's type and series number. The computer will also check to see if there are sensors that fall outside the specified tolerances.

This automatic measurement system can also be used during sensor research phases to establish, for instance, the optimal measurement frequency. To that end, by measuring at a range of different frequencies (see Figure 20.18) the system can be made to produce a "three-dimensional" plot.

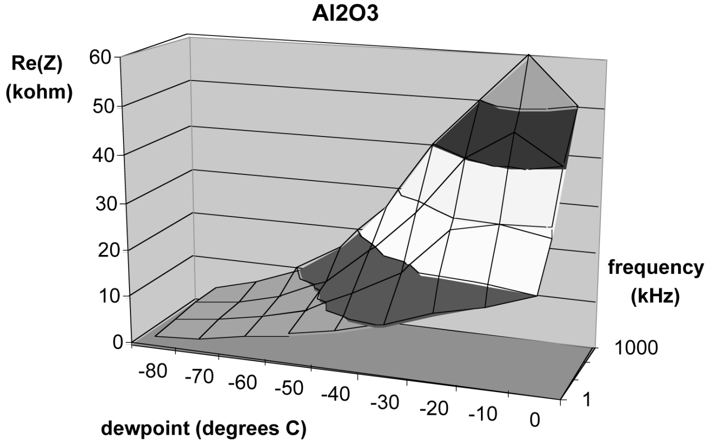


Figure 20.18. An example of a set of characteristics for the capacitive humidity sensor shown in Figure 20.17, obtainable by means of computerized measuring. Here we see the real part of sensor impedance plotted as a function of frequency and humidity.

These pictures give a direct insight into the behavior of the sensor in relation to several parameters. It is also possible to plot a number of other parameters – all versus frequency – such as: the real and imaginary parts of impedance, the modulus and the argument, all versus frequency. As sensor behavior is hard to predict it is possible that other output parameters may be more suitable for measuring humidity. Once that information has been collected the optimal output parameter and measurement frequency can be determined.

On the basis of the same measurement system much more information can be retrieved, such as information on the sensor's temperature sensitivity, its response time and the effect that process parameters have on the characteristic properties of the sensor. The purchase of such expensive automatic measuring instruments may be justified if the measurement sequences being dealt with are ones that have to be repeated many times.

20.2.3 Virtual instruments

Having now introduced the PC as part of a measuring instrument we shall now go on to consider some additional features of this approach. The PC substantially increases the flexibility of the measurement system. Huge amounts of data can be stored on disk for complex, off-line data processing or just for archiving purposes. With the appropriate software it becomes possible to make a statistical analysis of the measurement data, to analyze particular quantity trends and to establish the correlation between the two or more quantities investigated. Graphic representation facilitates the quick evaluation of single experiments and the representation of complex processes on a single display unit, thus eliminating the need to read individual displays for each measurand.

The functionality of the PC can be further extended by introducing a function generator (for specific test signals), a filter (in the digital domain), a controller (that uses the acquisition card's output channels) and many more devices. With these extensions the PC incorporates several instruments in a traditional measurement (and control) system. Moreover, since the PC is a freely programmable device, measurement structures and procedures can be set up and controlled in an arbitrary way.

A further step in the direction of increasing flexibility is the applying of object-oriented programming to the design of a measurement system. Since many functions have already been implemented as PC software, the configuring of a measurement system with graphic symbols (icons) constitutes a small step in the direction of virtual instrumentation.

Classical computer programming is based on a declarative approach in which commands, according to some languages, are written into a program file. Former programmable measurement instruments used this type of programming.

This programming technique may be termed command-driven. However, a measurement system is data driven: the data is or should be immediately processed as soon as it is produced. Several vendors have identified "visual languages", like for example Simulink which is part of MATLAB, HP-Vee and Labview. Icons representing specific actions that have to be performed are placed on a “worksheet”. These icons also have input and output terminals. By graphically connecting the output terminal of one icon to the input terminal of another one is able to define a data-stream. In that way, entire networks of instruments and their connections can be simulated. Real or simulated data can be connected to the inputs of the icons as though they are real instruments. After starting the program and completing a logic check of the network, the icons perform similar actions with the input data and the resulting output data appears at the icon's output terminal.

Although the program is data-driven, the PC architecture remains command driven. This means that all the icon's input terminals have to be continuously polled. Moreover, PCs contain processors that execute all tasks consecutively: no two actions can be executed at the same time so this limits the possibilities when it comes to the carrying out of real-time tasks. Dedicated PC-boards with their own processing hardware considerably alleviate this problem.

SUMMARY

Electronic measurement instruments

- A multimeter is an instrument that is used to measure various electrical quantities. Usually these are: voltage, current and resistance.
- With an oscilloscope one or more electrical signals can be visually represented as a function of time. Periodic signals can be represented continuously as stable images. The horizontal scale (time base) and the vertical scale (signal amplitude) are adjustable.
- Signal generators produce electrical signals such as, according to the type of instrument: DC signals, sine waves, triangular waves, ramp and pulse signals and noise. Function generators produce various periodical signals, sometimes simultaneously. The output amplitude and the frequency are adjustable over a wide range.
- A frequency counter is suited to the measurement of the frequency of periodic signals, to their period time or to the number of pulses or zero crossings.
- With a spectrum analyzer, the frequency spectrum of an electrical signal is presented in visual form for an adjustable range of frequencies.
- Network analyzers make it possible to investigate the impedance of two-port networks over a wide frequency range. Depending on the type of instrument, Bode plots (amplitude characteristic and phase characteristic) can be visually displayed.
- An impedance analyzer is suitable for the measurement of the various impedance parameters of a (passive) two-pole network as a function of frequency.

Computer-based measurement instruments

- Computers speed up measurement time and facilitate data presentation. The computer is used to control the connected measurement instruments and to store and present the measurement data.
- The various instruments in an automatic measuring system are connected via a bus structure. A widely-used standardized bus is the IEC 625 bus (or the IEEE 488 bus).
- The IEC instrumentation bus is based on asynchronous data transport. To stay independent of the processing speed of the connected instruments, the data transfer is controlled by a handshake procedure.
- In virtual instruments the PC takes over various functions from traditional measurement instruments, all those functions are replicated in software. A measurement system is configured by means of object-oriented programming, and the measurements are conducted automatically according to a preprogrammed sequence. A virtual instrument is nevertheless a real instrument when it comes to the matter of measuring real physical quantities.

EXERCISES

Electronic measurement instruments

20.1 Three volt meters are calibrated for rms values. Meter A is a true rms meter; meter B measures the average of the modulus (double-sided rectified signal) and meter C measures the average of the signal clamped with its negative peaks to zero (see Section 9.2.3). The three meters are calibrated correctly for sinusoidal input signals. The meters are used to measure three different signals (see the table below), all symmetrical, with an average of zero and a peak value of 10 V. What do these instruments indicate?

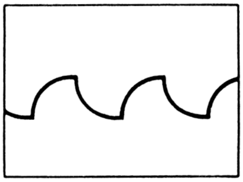
input voltage:	meter A	meter B	meter C
sine wave			
square wave			
triangular wave			

20.2 Why do some multimeters have four terminals for a resistance measurement instead of only two.

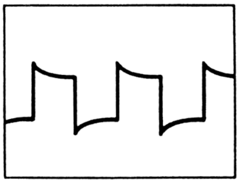
20.3 Describe the principles of the alternate mode and the chopped mode in a multi-channel oscilloscope. What is an easy way of checking to see whether an oscilloscope is in the alternate mode?

20.4 Draw a picture of the amplitude transfer characteristic V_o/V_i in Figures 20.6a and b (with and without probe measurements). The components have the following values: R_g = 250 kΩ; R_i = 1 MΩ; C_i = 26 pF; R = 9 MΩ; C = 24 pF; C_k = 190 pF. The resistance R_g may be ignored compared to R.

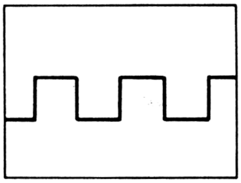
20.5 The following figure shows three possible pictures of the square test signal on the oscilloscope screen during probe adjustment. Which of the pictures corresponds to correct adjustment and why?



a



b



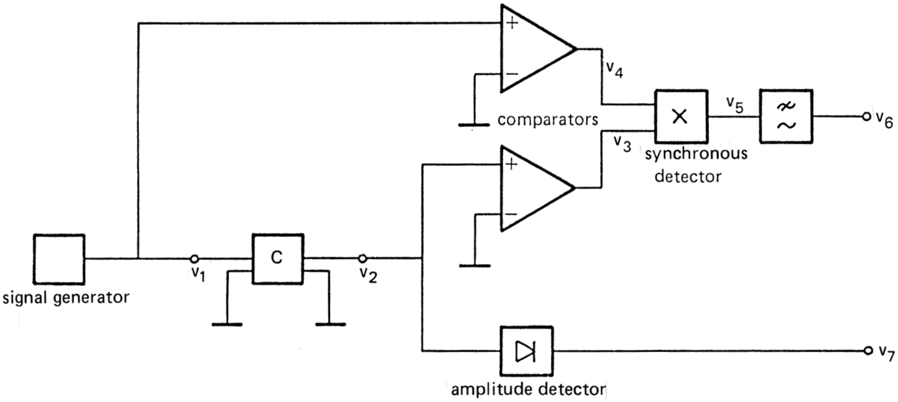
c

20.6 The oscilloscope can produce what are known as Lissajous figures. To that end, a sinusoidal voltage $y(t) = \sin\omega t$ is connected to the Y-input of the oscilloscope, and a voltage $x(t) = \sin(\omega t + \varphi)$ to the X-input (Figure 20.4). Suppose $\omega = 825$ rad/s. Plot the picture on the oscilloscope for the following values of φ: a. 0°; b. 90°; c. 180°; d. 270°; e. 300°.

20.7 Digital noise generators are based on digital, logical circuits. What does the output of such a generator look like?

20.8 The oscillator in the next given network analyzer produces a sine wave signal $v_1 = A\sin\omega t$. The amplitude of the square wave output signals of the comparators is

B; the transfer of the network is $C = |C|e^{j\varphi}$. Make an amplitude time diagram of all other signals in this system.



Computer-based measurement instruments

20.9 Imagine a computer-based measurement system based on the IEC 625 bus. This system must be used to plot the Bode plot for the transfer of arbitrary two-port networks.

The requirements are: a plotter resolution of 0.4 mm for the lengths of the axes, 40 cm for the frequency axis and 20 cm for both the amplitude and phase axes, a frequency range of 10 Hz to 100 kHz, an amplitude range of 40 dB, a phase range of 2π rad.

a. Plot the structure of such a measurement system;

b. Find the relative frequency steps;

c. Calculate the required amplitude detector resolution (in dB);

d. Calculate the required phase meter resolution (in radians or degrees).

20.10 Explain the instrument functions “listen” and “talk”.

20.11 Explain the handshake procedure and indicate also what implications this has for data transmission speed?

20.12 Discuss the main aspects of “virtual instruments”.

21 Measurement uncertainty

21.1 Measurement uncertainty described

It is impossible to ever be completely certain about the value of any physical quantity since a degree of uncertainty, sometimes also termed inaccuracy or error, is inherent in any measuring process. When presented, measurement results should provide some indication of the inaccuracy of the measured values. Without such indications the measurement data is of little significance. The simplest way to indicate the degree of inaccuracy is by giving the number of significant digits. As a rule, it is only the last and least significant digit that is inaccurate. In such cases the inaccuracy interval is approximately half the unit of the last digit, unless otherwise specified.

Example 21.1
The given measured value of a voltage is 10 V. The true value lies between 9.5 and 10.5 V. If the voltage that is given is 10.0 V, then the actual value will lie between 9.95 and 10.05 V. The notation for a more precise tolerance interval (if it is possible to specify that) will, for instance, be 10.00 ± 0.02 V. Notations like 10 ± 0.02 V and 10.000 ± 0.02 V should never, however, be used.

21.1.1 Types of uncertainty

Throughout this book we distinguish between two types of measurement errors, systematic errors and random errors. Human errors or mistakes, caused by miscalculating or incorrectly reading the displays or the settings of instruments are disregarded. Errors of that kind can be avoided if one works conscientiously. A systematic error is an error that remains the same when measuring is repeated under identical conditions. Possible causes of such errors are wrong calibration, offset and impedance mismatch (Section 5.1.3).

Example 21.2
The short-circuit current of a voltage source is measured using a current meter with input resistance R_M. In order to establish the true value of the short-circuit current the

measured value must be multiplied by $(R_g + R_M)/R_g$. *If no correction is made for loading error the measurement result will show a systematic error.*

Systematic errors can be traced or eliminated by using other instruments to repeat the measurement, by changing the measurement method, by recalibrating, by carefully inspecting both the measurement system used to analyze and the measurement object (i.e. by means of correct modeling) or by introducing correction factors in the way illustrated in the preceding example.

When measurements are repeated the random errors will have different values. Possible influencing factors will be interference from outside or noise generated by the measurement system or measurement object. As it is impossible to predict the value of the error it has to be described in terms of probability (Section 2.2.4). In most cases random errors exhibit normal distribution or Gaussian distribution (Figure 21.1).

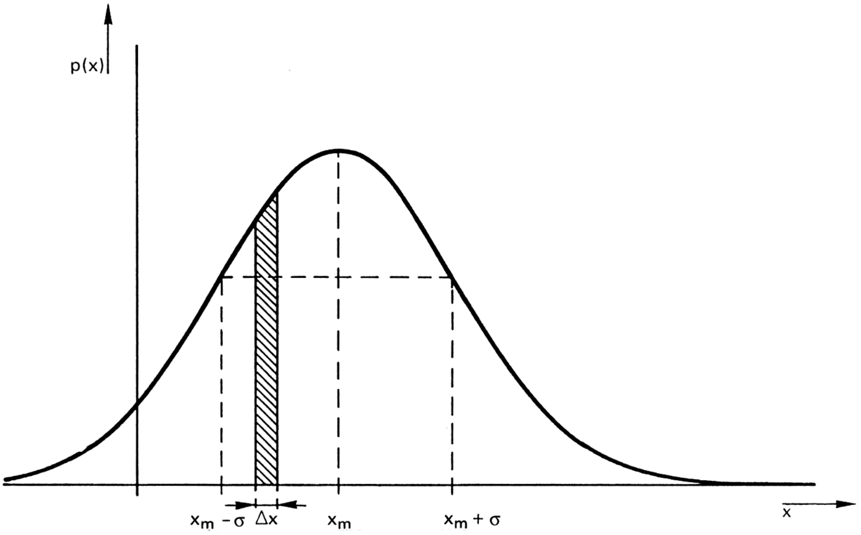


Figure 21.1. The probability density function $p(x)$ of a stochastic parameter with normal or Gaussian distribution. The shaded area is a gauge of the probability of the measurement value actually being within the interval Δx.

The probability of the value being within the interval $[x, x + \Delta x]$ is equivalent to

$$P(x, x + \Delta x) = \int_{x}^{x+\Delta x} p(x)dx \tag{21.1}$$

(see the shaded area in Figure 21.1). The maximum occurrence probability lies with the mean or average value. This average:

$$x_m = \int_{-\infty}^{\infty} p(x)xdx \tag{21.2}$$

constitutes the best estimation of the measurement value. The actual values are scattered around the average. An estimation of the dispersion of the measurement values is the variance which is defined as:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - x_m)^2 p(x) dx \tag{21.3}$$

The variance σ^2 is nothing other than the average of the squared deviation from the mean value x_m. The square root σ is termed the deviation or the standard deviation.
When the density function probability is known, it is possible to calculate the probability of the true measurement value actually being outside the interval $[x, x + \sigma]$. Table 21.1 shows this probability for various intervals around the normal distribution function mean.
It may be viewed as the reliability of the measurement value estimation. Since the probability of greater error is never zero it is impossible to specify the maximum value of a random error. Sometimes, though, the maximum error is specified anyway. In such cases it is usually the 3σ error that is intended. The probability of a true value being within the $\pm 3\sigma$ interval is 0.9972 (Table 21.1).

Table 21.1. The probability of the true value having a deviation that is more than $n\sigma$ from the average of a normally distributed quantity.

Interval	probability of being outside the interval
$X_m \pm 0.6745\sigma$	0.5000
$X_m \pm 0\sigma$	0.3172
$X_m \pm 2\sigma$	0.0454
$X_m \pm 3\sigma$	0.0028

Up until now we have simply considered the measurement quantity as a continuous variable. Obviously the amount of measurement data is limited and the intervals Δx are of a finite width. In such cases the probability density function is actually a histogram (Figure 21.2).
For a finite number of measurement values the mean value is given as

$$x_m = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{21.4}$$

in which n is the number of measurements and x_i the result of the i-th measurement. The variance is

$$Var(x) = \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - x_m)^2 \tag{21.5}$$

Here, too, the deviation or standard deviation is the square root of the variance.

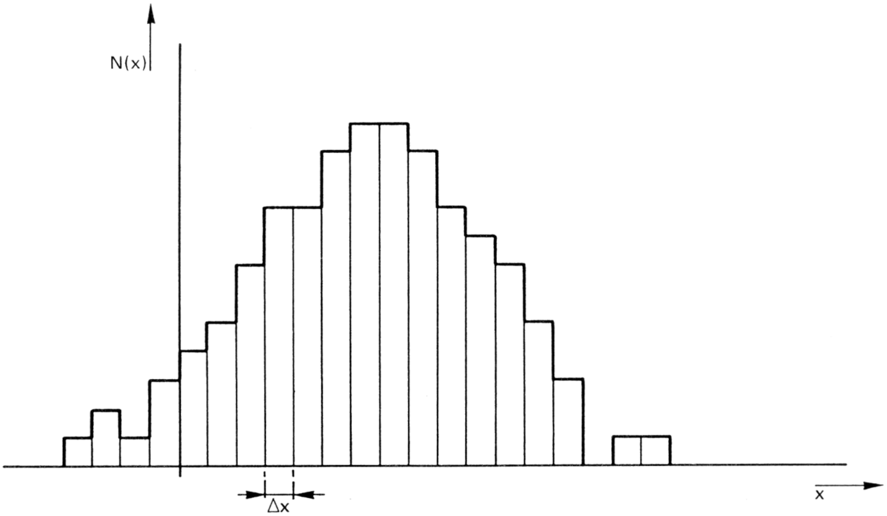


Figure 21.2. A histogram reflecting the number of measurements *N*(*x*) with results that lie within the Δ*x* interval as a function of *x*.

21.1.2 Error propagation

Often the value of a physical quantity or parameter emerges from the measurement result calculations made for several other quantities. The errors emanating from those kinds of measurements will affect the accuracy of the calculated value. In this section we shall present several ways of finding the error in the final measurement result.

In Section 1.2 we introduced absolute inaccuracy and relative inaccuracy which are the two ways of specifying inaccuracy. An absolute error is the difference Δx between the true value x_o and the measured value x: $\Delta x = |x - x_o|$ while relative error is defined as $|\Delta x/x| = |(x - x_o)/x$. Both the absolute and relative errors are positive, unless explicitly otherwise specified. The true value should lie within the interval $[x - \Delta x, x + \Delta x]$. The absolute error of a resistor with a resistance value of 80 ± 2 Ω is, for instance, 2 Ω, whereas its relative error is 1/40 or 2.5%. Evidently, the actual deviation might be less.

If we assume that quantity z is a function of several measurement parameters: $z = f(a,b,c,...)$, each with an absolute inaccuracy Δa, Δb, … then the maximum error in z will be found to be:

$$\Delta z = \left|\frac{\partial f(a,b,c,\ldots)}{\partial a}\right|\Delta a + \left|\frac{\partial f(a,b,c,\ldots)}{\partial b}\right|\Delta b + \left|\frac{\partial f(a,b,c,\ldots)}{\partial c}\right|\Delta c + \ldots \qquad (21.6)$$

The total error is the sum of the absolute errors of the individual measurement errors: it is the maximum absolute error. The true error is less because some of the errors are positive while others are negative, so they might partly cancel each other out. The total relative error can be derived from the total absolute error expression:

$$\left|\frac{\Delta z}{z}\right| = \left|\frac{\partial f(a,b,c,\ldots)}{\partial a}\cdot\frac{a}{f(a,b,c,\ldots)}\cdot\frac{\Delta a}{a}\right| + \left|\frac{\partial f(a,b,c,\ldots)}{\partial b}\cdot\frac{b}{f(a,b,c,\ldots)}\cdot\frac{\Delta b}{b}\right| + \left|\frac{\partial f(a,b,c,\ldots)}{\partial c}\cdot\frac{c}{f(a,b,c,\ldots)}\cdot\frac{\Delta c}{c}\right| + \ldots \quad (21.7)$$

Some fixed guidelines for the calculating of the absolute and relative errors can be derived from these two expressions as given below without further proof.
Suppose $z = f(a,b)$, with $z = z_o \pm \Delta z$, $a = a_o \pm \Delta a$ and $b = b_o \pm \Delta b$.

- for addition ($z = a + b$) and subtraction ($z = a - b$) the absolute errors involve adding: $|\Delta z| = |\Delta a| + |\Delta b|$;
- for multiplication ($z = a.b$) and division ($z = a/b$) the relative errors involve adding: $|\Delta z/z| = |\Delta a/a| + |\Delta b/b|$;
- in the case of power functions ($z = a^n$) the relative error is multiplied by the modulus of the exponent n: $|\Delta z/z| = |n|\cdot|\Delta a/a|$.

Example 21.3
The measured current through a resistance of 50 ± 1 Ω is 0.40 ± 0.02 A. The nominal voltage across the resistance is V = IR = 20 V, and the relative error is equal to 1/50 + 0.02/0.40 = 2% + 5% = 7%. The absolute error may therefore be said to equal ±1.4 V.

The dissipated power is $P = I^2R = 8$ W with a relative error of 2×5% + 2% = 12%. The calculated power is given as 8 ± 1 W.

It is useless to specify the maximum absolute error when all the errors are random. In such cases we must therefore specify the mean value and the variance. The mean value of $z = f(a,b,\ldots)$ is $z_m = f(a_m,b_m,c_m,\ldots)$ and the variance is:

$$\sigma_z^2 = \left\{\frac{\partial f(a,b,c,\ldots)}{\partial a}\right\}^2 \sigma_a^2 + \left\{\frac{\partial f(a,b,c,\ldots)}{\partial b}\right\}^2 \sigma_b^2 + \left\{\frac{\partial f(a,b,c,\ldots)}{\partial c}\right\}^2 \sigma_c^2 \ldots \quad (21.8)$$

In the last expression for σ_z^2 the mean values of the variables a, b, etc. need to be substituted. On the basis of these expressions it is possible to calculate the mean and the variance of the final measurement result with random errors. These rules also apply to quantities with a distribution function that deviates from normal distribution.

21.2 Measurement interference

One of the leading causes of measurement errors is interference or noise instigated by the interaction of the environment with the measurement system. Often these errors can be avoided by simply taking proper measures. Some of the causes of interference and their remedies will be discussed in this final section.

21.2.1 Causes of interference

Errors that are due to interference are usually random errors, which can easily be recognized as such by their fluctuating character. We shall now briefly review the main causes of interference in electronic measurement systems.

- The capacitance of a cable connecting the measurement system with the measurement object may show fluctuations that derive from mechanical vibrations (Figure 21.3).
  The input voltage V_i of the measurement system equals

$$V_i = \frac{R_i}{R_i + R_g + j\omega R_g R_i C_k} V_g \tag{21.9}$$

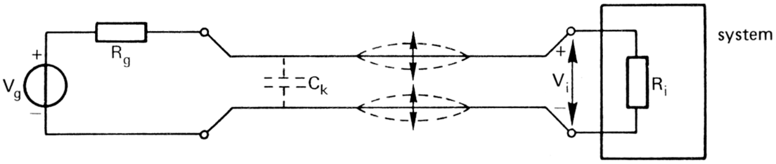


Figure 21.3. Capacitance fluctuations that are due to mechanical vibrations.

so when, at constant source voltage, cable capacitance C_k changes, the input voltage varies accordingly. The effect can be reduced or even avoided by using high quality cables that are properly fixed.

- The signal transfer of a measurement system may vary according to temperature and may be due to the temperature coefficient of the components (Section 1.1); small DC error signals will be generated because of the thermoelectric effect (Section 7.2.4) which will be particularly apparent when the input connections (possibly consisting of different materials) are subjected to temperature gradients.
- Electric signals can be induced into the measurement system capacitively via stray capacitances like for instance the capacitance between the system input terminals and the mains (Figure 21.4).

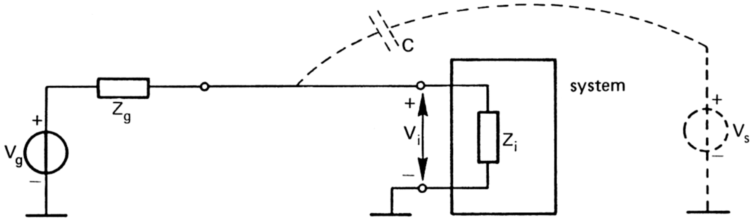


Figure 21.4. Capacitive signal injection via stray capacitances.

It is not only mains voltage (50 Hz) but also other signals originating from, for instance, fuel engines (ignition), electric motors (commutator switching) and thyristor circuits that give rise to capacitively injected interference signals. These types of systems produce high voltage peaks that can easily enter the measurement system, even if the capacitance is very small. Using the model seen in Figure 21.4 the input voltage that is solely due to the error signal is found to be

$$V_{i,s} = \frac{Z_g // Z_i}{Z_g // Z_i + 1/j\omega C} V_s \tag{21.10}$$

Obviously in situations where both Z_i and Z_g have high values the error input signal can actually be very high. The effect produced is known as hum after the noise produced in audio instruments (where similar effects may well occur).

- It may also be because of magnetic induction that error signals are able to enter the measurement system (see Section 7.1.3). The situation described is depicted in Figure 21.5.

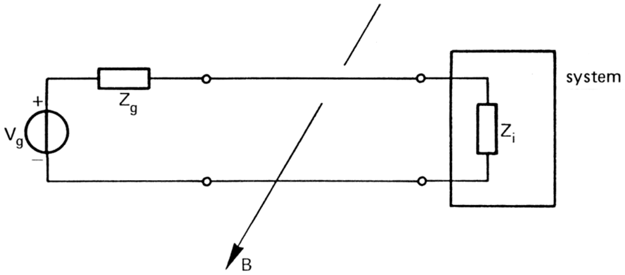


Figure 21.5. Inductive error signal injection due to varying magnetic field loop intersection.

Stray magnetic fields are produced when currents flow through conductors (e.g. mains supplies) and other power instruments (e.g. welding transformers). The induced error voltage equals $v_s = -A\mathrm{d}B/\mathrm{d}t, A$ being the surface area of the conductor loop and B the magnetic induction (Section 7.1.3). The error increases as the frequency and loop area increase.

- For safety reasons the metal casing of a measurement instrument is always earthed. Such measurement systems can sometimes be the cause of error injection. Many systems have the instrument ground (i.e. the circuit's earth) connected to the casing as well and thus also to the safety earth. The interconnecting of such instruments may give rise to earth loops (Figure 21.6a).

  If the magnetic field is variable that will induce a voltage in the ground loop conductor which will, in turn, result in a ground loop current. As ground conductors have a non-zero resistance (of generally around 0.1 Ω) an error voltage will be generated that is in series with the measurement system input. The same happens when the systems are connected both to the signal and to the return lines (Figure 21.6b). Such loops are sometimes unavoidable, in particular when using coaxial cables for the signals.

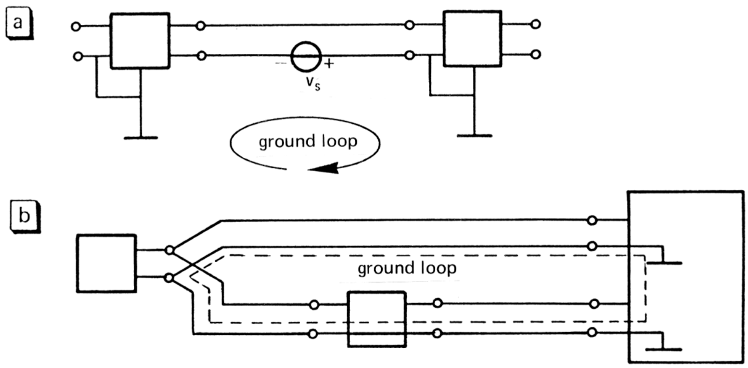


Figure 21.6. Two examples of ground loops, (a) where instruments are interconnected via the safety earthing, (b) where instruments interconnect with signal pairs.

- Error signals (Figure 21.7) may also derive from shared ground connections. The current flowing through the common ground (or return line) may be considerable if other instruments are using the same ground as a return line. The error voltage is $I_g R_{ground}$, which is in series with the input signal source.

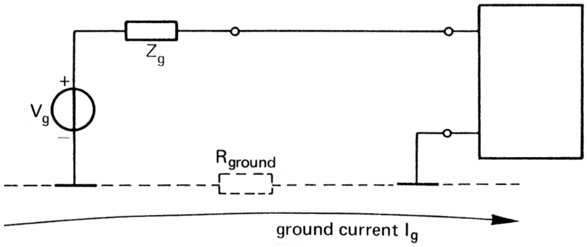


Figure 21.7. Due to the unknown ground currents, common ground connections may give rise to error signals.

21.2.2 Remedies

In this section an overview will be given of some of the measures that can be taken to reduce or altogether eliminate error signal interference in measurement signals. The possible general measures that can be taken are these:

- remove the source of error
- isolate the measurement system error source
- compensate
- correct
- separate

- The technique of simply removing the source of the error can, where this is indeed possible, prove to be a very effective tool. However, it can only be done if the source can be switched off without this causing any problems. Increasing the distance between the error source and the measurement system can be very effective.

- Either the measurement system or the error source (or possibly even both) can be isolated to prevent interference. Usually the simplest solution is to isolate the measurement system. We shall now give some examples of how this can be done.

Example 21.4

By placing part of the measurement system in a temperature-stabilized compartment the effect that temperature variation has on the measurement can be eliminated. This is the method often applied to accurate and stable signal generators and reference sources where the critical components (i.e. the crystal, the Zener diode) are kept at a constant temperature.

Example 21.5

Capacitive error signal injection can be drastically reduced by shielding (Figure 21.8). The shield must be a good conductor and firmly connected to the ground. The induced signals will subsequently flow to the ground and will not be able to reach the measurement system input. Likewise, the instrument can be shielded from magnetic error signals as well by using magnetic shielding materials (i.e. metals with very high permeability and therefore low magnetic resistance).

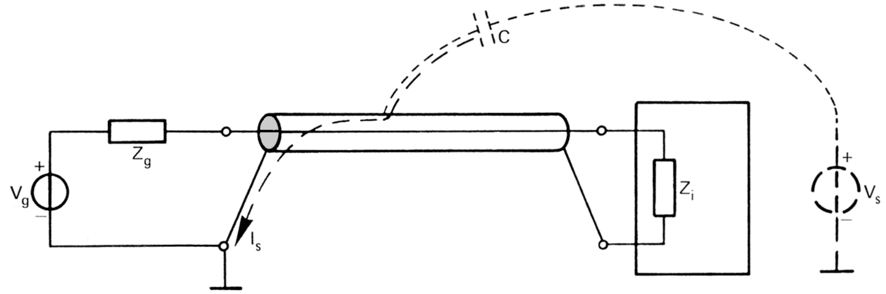


Figure 21.8. A grounded shield will reduce capacitive error signal injection between the source and the measurement circuit.

- Compensation is the technique most widely used in electronics to get rid of unwanted signals and sensitivity. One very effective compensation method is the following one which is based on two or more simultaneous measurements in which, for instance, two sensors are arranged in differential mode: one sensor signal increases and the other decreases as the measurand increases. The sensitivity to interference signals is the same in both sensors. The error appears as a common mode signal that can be eliminated by means of a differential amplifier with high CMRR (Section 1.2). The compensation method can be illustrated using the following two examples.

Example 21.6

A Wheatstone bridge (Figure 21.9; see also Example 17.2) contains two strain gauges (Section 7.2.1.3). One is sensitive to compressive strain and the other to tensile strain but they are both equally sensitive to temperature change.

If the temperature of the strain gauges is equal (and if the temperature coefficients are equal) then full temperature compensation will be achieved.

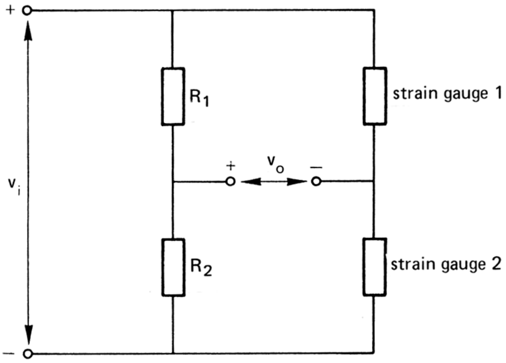


Figure 21.9. A Wheatstone bridge with two strain gauges for the elimination of temperature sensitivity. One of the gauges is sensitive to strain but both are equally sensitive to temperature.

Example 21.7
Interference due to magnetic induction in ground loops which may or may not be grounded is reduced significantly if the conductors are twisted (Figure 21.10).
If two adjacent loops have equal surface areas and if the magnetic induction field is otherwise the same then the two induced voltages will also be equal: they will have opposite polarity due to the twisting and the induced voltages will be cancelled out.

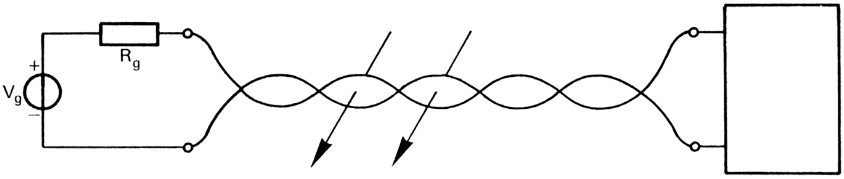


Figure 21.10. The twisting of conductor pairs will result in loop area reduction and thus also in induced voltage reduction.

- Measurement signal interference correction is usually based on the separate measurements of the interfering signal or signals. The temperature sensitivity of a sensor can, for instance, be corrected by measuring the temperature separately and by then going on to correct the measurement result on the basis of this data and the known temperature coefficient of the sensor.
- Measurement signal separation from interference signals is performed on the basis of frequency selective filtering. The method can only be applied when the frequency bands of the measurement signal and the error signal do not overlap. Signals that have relatively high frequencies can be freed from drift and other low-frequency signals by means of a simple high-pass filter. When the bands overlap the signal frequency may be converted to an adequate frequency band using modulation techniques (refer to Chapter 17).

Example 21.8
Environmental light (sunlight; 100 Hz lamp light) can easily interfere with an optical measurement system. In order to eliminate such interference signals the intensity of the optical source signal can be modulated using a relatively high frequency, for instance a few kHz. The receiver system converts the optical signals to electrical signals and then the interference signals are removed with a high-pass filter (Figure 21.11).

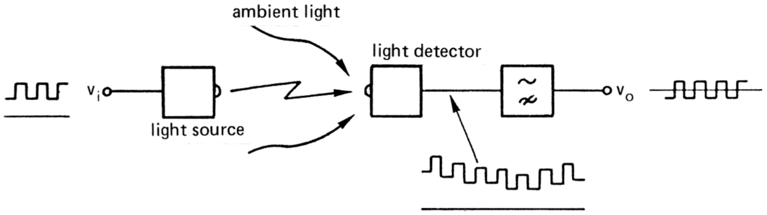


Figure 21.11. If a high frequency carrier is modulated this will allow unwanted low-frequency signals to be filtered out. Optical interference signals can thus be eliminated.

We shall now discuss two final ways of reducing possible measurement error. The first method involves eliminating the influence of cable impedance in systems where long interconnection cables are used to link the signal source (a sensor) with the measurement system. In such cases it is necessary to shield so as to reduce capacitive error signal injection. However, the grounded shield has a relatively large capacitance to ground, which is in parallel to the measurement system's input terminals. The cable impedance thus becomes part of the voltage transfer (Figure 21.12a).

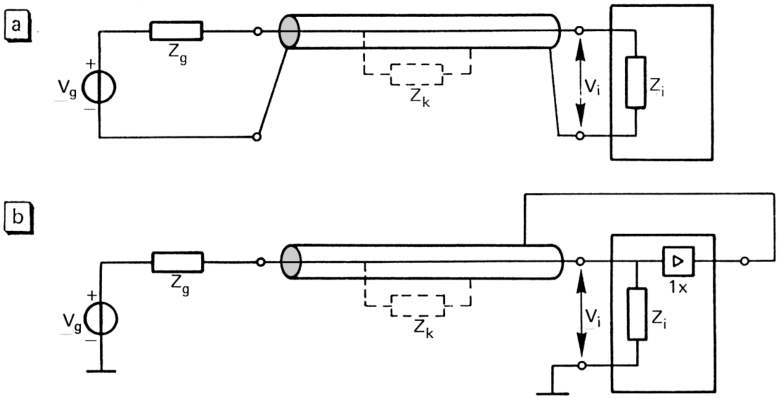


Figure 21.12. (a) Cable impedance Z_k is part of the signal transfer function V_i/V_g, b) active guarding eliminates the effect of cable impedance on signal transfer because then the cable voltage is zero.

Particularly when the impedance of the source is high, cable impedance will greatly affect the transfer. One way of reducing such an effect is by actively guarding: instead of grounding the shield it is connected to the output of a buffer amplifier, the input of which is equal to the signal voltage. The voltage across the cable is thus brought back to zero and no current will flow through the cable impedance. The shielding remains effective because the error signal to the shield flows into the buffer output without harming the input signal. Many instruments that are used to measure small currents or which have high input impedance have an extra guard connection to allow active guarding.

The second and last way of guarding against interference involves the elimination of unwanted ground currents (see Figure 21.7). The best remedy is to ground all instruments at a single point by creating a star configuration (Figure 21.13).
By grounding in this way no ground currents from other instruments or systems will be allowed to enter the measurement system.

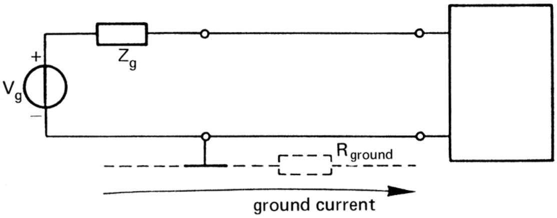


Figure 21.13. If there is a single-point ground connection this will prevent ground loop currents from entering the measurement system.

Unfortunately in practice it is possible for many error sources to be present simultaneously. It is especially the mains network surrounding the measurement system that causes a great deal of trouble when small signals are being measured. This is the reason why most instruments are equipped with metal casing and with shielded (coaxial) input cables. Nevertheless, shielding and other methods are seldom perfect. It is even conceivable that in eliminating one source of error one introduces another or makes the end-result worse because previously two different error signals might partly have been compensating each other. In such cases it is very tempting to renounce the new measures when such apparent deterioration is observed.

SUMMARY

Measurement uncertainty described

- The presentation of any measurement result must incorporate an indication of the inherent measurement inaccuracy, for instance by indicating the number of significant digits.
- Two possible types of errors are systematic errors and random errors. A systematic error will be constant in identical experiments while random errors can be described in terms of probability.
- The best way of estimating the true value of a measured quantity is by taking the mean of all the measured data. A measure of the deviation from the true value is the variance.
- The accuracy of any measurement result can be specified in terms of absolute and relative inaccuracy. To find the inaccuracy of a quantity that represents the sum (or

difference) of various measurement parameters, the absolute errors of those parameters must be totaled. In cases where one talks of the product or ratio between parameters the relative errors must be added, in cases of power function, the relative error of the parameter will be raised to the power $|n|$ and n will be the exponent.

Measurement interference

- Any measurement system is sensitive to interference or, to unwanted signals that may enter the system and reduce the accuracy of the measurement.
- The major causes of interference are mechanical vibration, changes in ambient temperature, capacitive and inductive error signal injection and errors derived from ground currents.
- The general ways to remedy interference are by: removing, isolating, compensating, correcting and separating. Ground errors can be reduced by introducing a single ground connection.

EXERCISES

Describing measurement uncertainty

21.1 The voltage difference V_{ab} between two points a and b is determined by measuring both voltages at a and b relative to ground: $V_{ab} = V_a - V_b$. The specified relative inaccuracy of the voltmeter is ±1%; the measured values are $V_a = 788$ mV and $V_b = 742$ mV.

a. Calculate the absolute and relative errors in V_{ab}.

b. Calculate these errors in a case where V_{ab} is measured directly between the two terminals of this voltmeter.

21.2 The measurement of a current is repeated five times. The measured values are: 0.62, 6.21. 6.23, 6.18 and 6.24 mA.

a. What type or types of errors might be involved: mistakes, systematic errors, random errors?

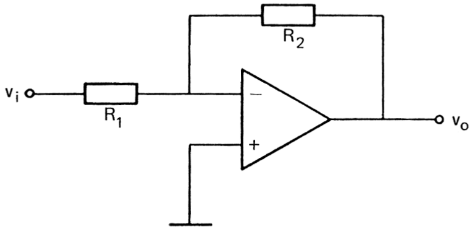
b. Find the best way to estimate the true value.

c. Think of a way of improving the accuracy of the measurement.

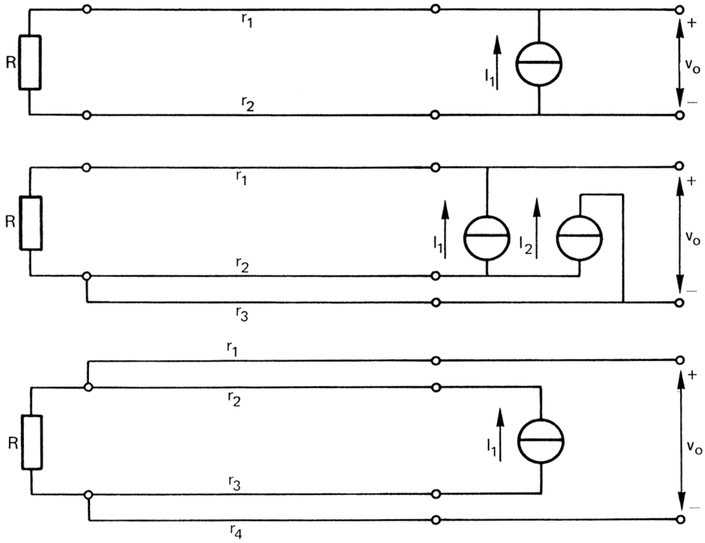
d. The experiment is repeated with a more precise current meter, the result is 6.25 ± 0.02 mA. What is the systematic error in the previous measurement?

21.3 A sensor voltage v_i from a sensor is amplified using the amplifier depicted below on the basis of the following specifications: $R_1 = 3.9$ kΩ ±0.5%; $R_2 = 100$ kΩ ±0.5%; $|V_{off}| < 1.5$ mV and $|I_{bias}| < 100$ nA. The sensor voltage appears to be $v_i = 0.2$ V ± 0.2%,

Calculate v_o and find the error in this voltage which is divided into an additive error (in mV) and a multiplicative error (in %). Find the tolerance margins of v_o.



21.4 The resistance value R of a resistor at some distance from the measurement system is measured in three different ways known respectively as the two-wire, the three-wire and the four-wire methods (see figure below).



In all cases the same measurement system is used and the following specifications: $I_1 = I_2 = 1$ mA ±0.5%, the inaccuracy of the voltmeter is ±0.5% ±0.5 mV, the resistance r of the wires lies between 0 and 3 Ω and the difference between the resistance values of the wires, Δr, is less than 1 Ω.

Calculate for each of the three methods the resistance R and the inaccuracy. The measured output voltages v_o are 75 mV, 70 mV and 71 mV.

Measurement interference

21.5 Imagine that a measurement system is subject to capacitive error signal injection from the mains as shown in the figure below.

Calculate the peak value of the error signal at the measurement system input point and (if applicable) the signal-to-noise ratio in each of the following cases:

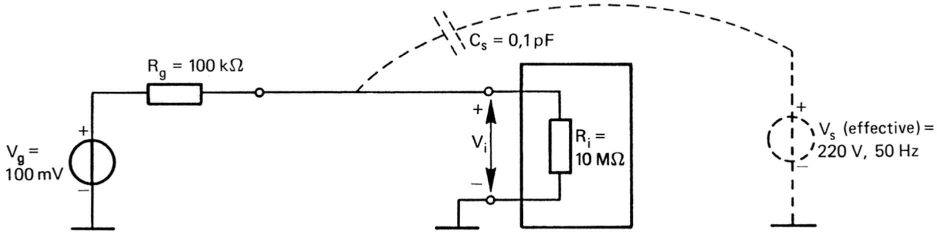
a. at disconnected signal source (floating input),

b. at connected signal source,

c. when there is a grounded shield that reduces capacitance C_s to 1% of the original value at disconnected source,

d. when the conditions are the same as in c but the source is connected,

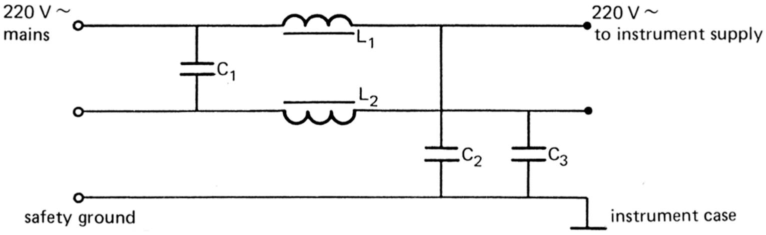
e. what would be the reduction of the interference signal if the shield were not grounded?


$C_s = 0,1$ pF
$R_g = 100$ kΩ
$V_g =$ 100 mV
+
−
+
V_i
−
$R_i =$ 10 MΩ
+
V_s (effective) = 220 V, 50 Hz
−

21.6 One way of reducing the errors caused by capacitive or inductive injection between the mains and the measurement system via the power supply transformer is by applying a mains filter in the way shown in the next figure.

a. Explain the method.

b. Explain what happens when the system is connected to the mains without the safety ground terminal being connected.


220 V ~ mains
L_1
C_1
L_2
C_2
C_3
220 V ~ to instrument supply
safety ground
instrument case

Appendix

A.1 Notation

A.1.1 Symbols

In this book we distinguish between quantities (and parameters) written in capitals and in lower-case. In general, quantities and parameters are written in lower-case letters except:

- static quantities;
- complex quantities;
- digital numbers.

Static quantities are: DC-signals, bias currents (I_{bias}), offset voltages (V_{off}), offset currents (I_{off}) and reference voltages (V_{ref}). Quantities for the biasing of electronic circuits belong also to this category.

The imaginary unit, in mathematical notation i, is written in electrical engineering as j, to avoid confusion with the symbol i for currents, hence $j = \sqrt{(-1)}$.

Digital words are quantities consisting of one or more bits, representing a binary number, for example A = $a_3a_2a_1a_0$ = 1101; this can be a binary numerical value ($1101_2 = 13_{10}$), the number of an address or the code for an instruction.

A current is positive in the direction of the arrow; the positive polarity of a voltage is indicated with a + sign. A voltage denoted with two indices indicates the voltage difference between the corrseponding points; the first point is positive with respect to the other. For example: V_{AB} is a voltage difference between the points A and B, where A is positive with respect to B (Figure A.1).

Consequently, $V_{AB} = -V_{BA}$. A voltage notation with only one index is the voltage difference between the point indicated by the index and a reference point, usually denoted as ground. The indexed point is positive. Figures A.1a and b are different notations for identical voltages and polarities.

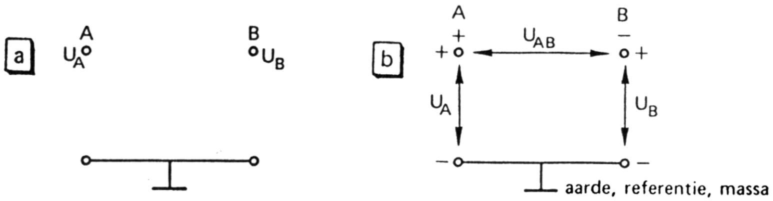


Figure A.1. Notation for voltages. (a) A voltage with one index (V_A, V_B) is the voltage difference as indicated in (b); for both figures is $V_{AB} = V_A - V_B = -V_{BA}$..

Needless to mention that these notations refer to the mathematical directions and polarities, and not to the physical values. A current flowing into a particular direction can be notated as a positive value (the direction of the arrow coincides with the physical direction) as well as a negative value (the arrow is opposite the physical direction).

A.1.2 Decimal prefixes

Table A.1. The prefixes for decimal multiples and submultiples.

factor	symbol	prefix	factor	symbol	prefix
10^{24}	Y	yotta	10^{-1}	deci	d
10^{21}	Z	zetta	10^{-2}	centi	c
10^{18}	E	exa	10^{-3}	milli	m
10^{15}	P	peta	10^{-6}	micro	µ
10^{12}	T	tera	10^{-9}	nano	n
10^{9}	G	giga	10^{-12}	pico	p
10^{6}	M	mega	10^{-15}	femto	f
10^{3}	k	kilo	10^{-18}	atto	a
10^{2}	h	hecto	10^{-21}	zepto	z
10^{1}	da	deka	10^{-24}	yocto	y

A.1.3 SI-units

In this book we use exclusively the notations and units according to the SI-system (Système International des Unités). We distinguish base quantities and derived quantities. The base SI-quantities are listed in Table A.2.

Table A.2. The base and supplementary SI quantities with corresponding units.

base quantity	SI unit	
length	m	metre
mass:kilogram	kg	kilogram
time	s	second
electric current	A	ampere;
thermodynamic temperature	K	kelvin
luminous intensity	cd	candela
amount of substance	mol	mole
plane angle	rad	radian
solid angle	sr	steradian

The definitions of the base and supplementary units are as follows:

> The **meter** is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second [17th CGPM (1983), Res.1].
>
> The **kilogram** is the unit of mass; it is equal to the mass of the international prototype of the kilogram [1st CGPM (1889)].
>
> The **second** is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom [13th CGPM (1967), Res.1].
>
> The **ampere** is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2 x 10^{-7} newton per meter of length [9th CGPM (1948)].
>
> The **kelvin**, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water [13th CGPM (1967), Res.4].
>
> The **mole** is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12 [14th CGPM (1971), Res.3]. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.
>
> The **candela** is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540 x 10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian [16th CGPM (1979), Res.3].

It is not very practical to express all quantities in the base SI-units; the unit for electric resistance, for instance, would be expressed as $kg.m^2s^{-3}A^{-2}$. This is the reason for specific SI-units for a number of derived quantities (Table A.3).

Table A.3. Some derived quantities with the corresponding SI-units.

derived quantity	symbol	name	expression in terms of other SI units	expression in terms of SI base units
frequency	Hz	hertz		s^{-1}
force	N	newton		$kg\ m\ s^{-2}$
pressure	Pa	pascal	N/m^2	$kg\ m^{-1}\ s^{-2}$
energy	J	joule	N m	$kg\ m^2\ s^{-2}$
power	W	watt	J/s	$kg\ m^2\ s^{-3}$
charge	C	coulomb		A s
electric potential difference	V	volt	W/A	$kg\ m^2\ s^{-3}\ A^{-1}$
electrical resistance	Ω	ohm	V/A	$kg\ m^2\ s^{-3}\ A^{-2}$
electrical conductance	S	siemens	Ω^{-1}	$s^3\ A^2\ kg^{-1}\ m^{-2}$
electrical capacitance	F	farad	C/V	$s^4\ A^2\ kg^{-1}\ m^{-2}$
magnetic inductance	H	henry	Wb/A	$kg\ m^2\ s^{-2}\ A^{-2}$
magnetic flux	Wb	weber	V s =	$kg\ m^2\ s^{-2}\ A^{-1}$
magnetic induction	T	tesla	Wb/m^2	$kg\ s^{-2}\ A^{-1}$
luminous flux	lm	lumen		cd
illuminance	lx	lux		$cd\ m^{-2}$

A.1.4 Physical constants

Table A.4 containes the numerical values of the major physical constants utilized in electrical engineering.

Table A.4. Some physical constants.

c	speed of light	$(2.997925 \pm 0.000003) \cdot 10^{8}$	m/s
μ_0	permeability of vacuum	$4\pi \cdot 10^{-7}$	H/m
ε_0	permittivity of vacuum	$(8.85416 \pm 0.00003) \cdot 10^{-12}$	F/m
e, q	electron charge	$(1.60207 \pm 0.00007) \cdot 10^{-19}$	C
k	Bolzmann's constant	$(1.3804 \pm 0.0001) \cdot 10^{-23}$	J/K
h	Planck's constant	$(6.6252 \pm 0.0005) \cdot 10^{-34}$	J s
π		3.14159265....	
e		2.71828183...	

A.2 Examples of manufacturer's specifications

The next two sections show the complete specifications of two widely used integrated circuits. The first is an example of an analogue circuit, an operational amplifier; the second one is an example of a digital circuit, a JK-flipflop. Other integrated circuits are specified likewise; they can be found in the data handbooks of the manufacturers.

A.2.1 Specifications of the μA747 (an analogue circuit)

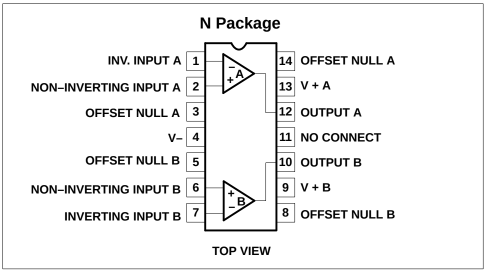
Philips Semiconductors Linear Products **Product specification**

Dual operational amplifier μA747C

DESCRIPTION

The 747 is a pair of high-performance monolithic operational amplifiers constructed on a single silicon chip. High common-mode voltage range and absence of “latch-up” make the 747 ideal for use as a voltage-follower. The high gain and wide range of operating voltage provides superior performance in integrator, summing amplifier, and general feedback applications. The 747 is short-circuit protected and requires no external components for frequency compensation. The internal 6dB/octave roll-off insures stability in closed-loop applications. For single amplifier performance, see μA741 data sheet.

PIN CONFIGURATION


N Package
INV. INPUT A 1
NON–INVERTING INPUT A 2
OFFSET NULL A 3
V– 4
OFFSET NULL B 5
NON–INVERTING INPUT B 6
INVERTING INPUT B 7
14 OFFSET NULL A
13 V + A
12 OUTPUT A
11 NO CONNECT
10 OUTPUT B
9 V + B
8 OFFSET NULL B
A
B
–
+
+
–
TOP VIEW

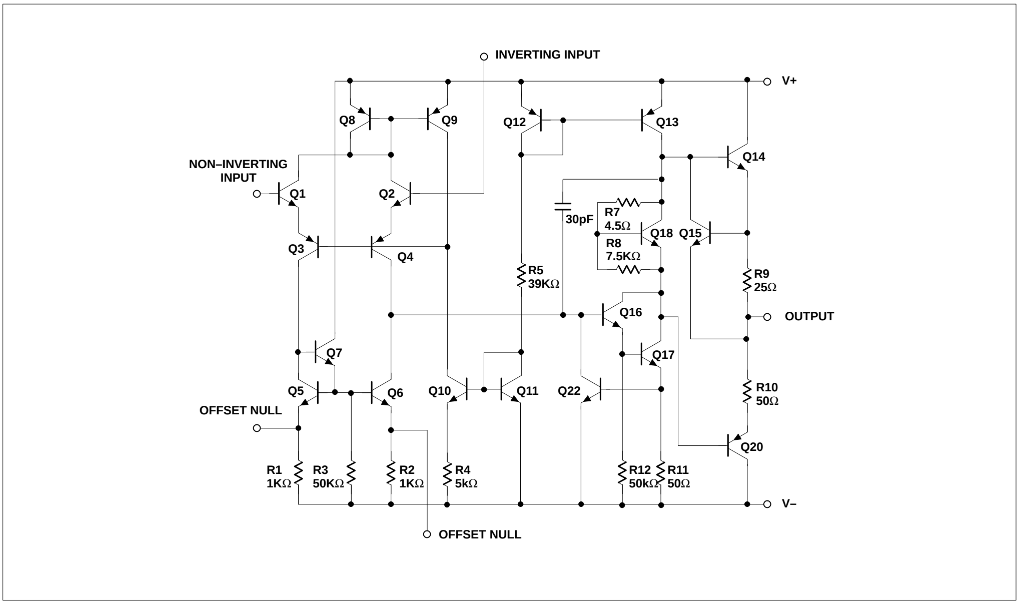
FEATURES

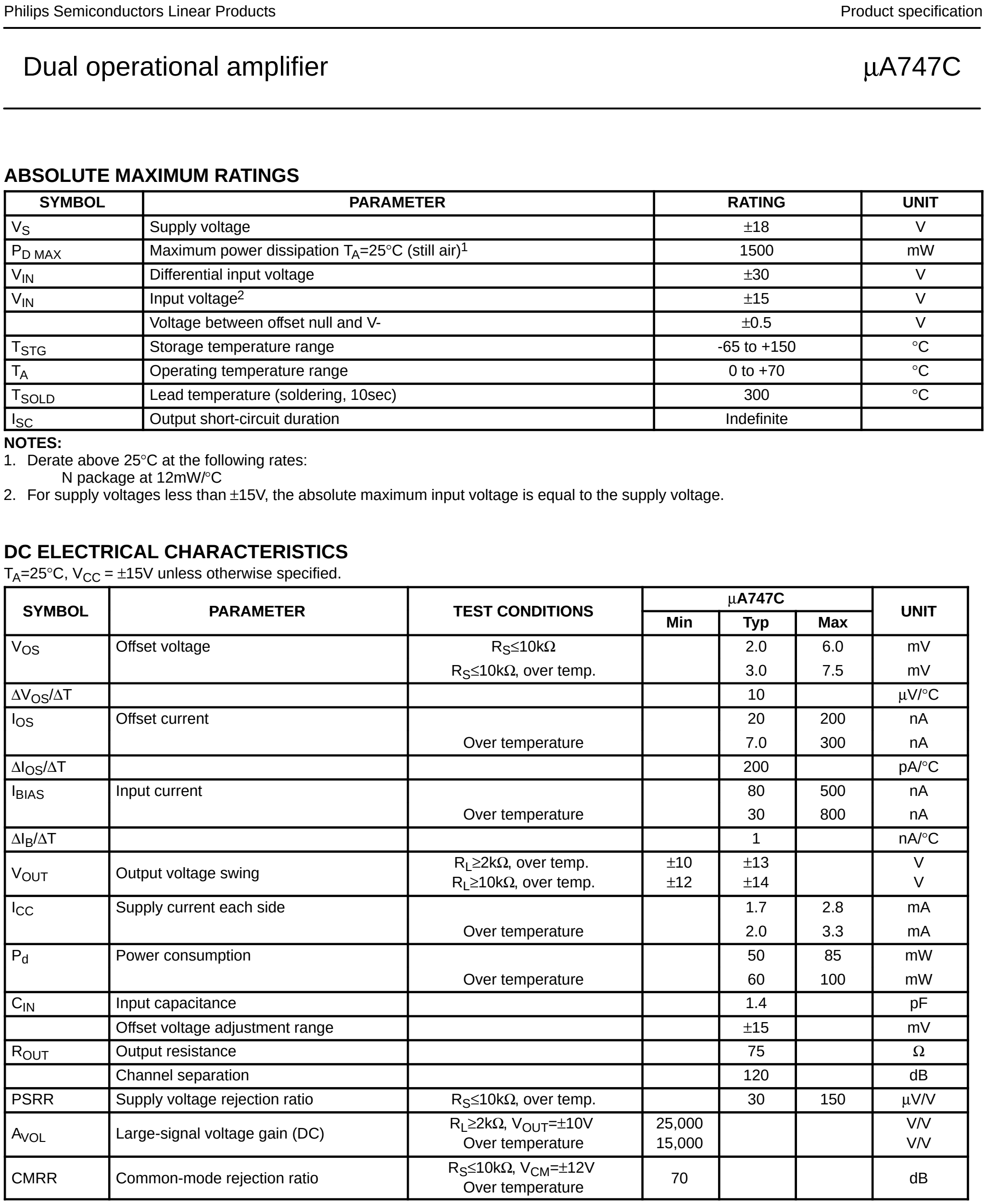
- No frequency compensation required
- Short-circuit protection
- Offset voltage null capability
- Large common-mode and differential voltage ranges
- Low power consumption
- No latch-up

ORDERING INFORMATION

DESCRIPTION	TEMPERATURE RANGE	ORDER CODE	DWG #
14-Pin Plastic DIP	0°C to 70°C	μA747CN	0405B

EQUIVALENT SCHEMATIC


INVERTING INPUT
V+
Q8
Q9
Q12
Q13
Q14
NON–INVERTING INPUT
Q1
Q2
30pF
R7 4.5Ω
Q18
Q15
Q3
Q4
R8 7.5KΩ
R5 39KΩ
R9 25Ω
Q16
OUTPUT
Q7
Q17
Q5
Q6
Q10
Q11
Q22
R10 50Ω
OFFSET NULL
Q20
R1 1KΩ
R3 50KΩ
R2 1KΩ
R4 5kΩ
R12 50kΩ
R11 50Ω
V–
OFFSET NULL



Philips Semiconductors Linear Products — Product specification

Dual operational amplifier — µA747C

ABSOLUTE MAXIMUM RATINGS

SYMBOL	PARAMETER	RATING	UNIT
V_S	Supply voltage	±18	V
$P_{D\ MAX}$	Maximum power dissipation T_A=25°C (still air)[1]	1500	mW
V_{IN}	Differential input voltage	±30	V
V_{IN}	Input voltage[2]	±15	V
	Voltage between offset null and V-	±0.5	V
T_{STG}	Storage temperature range	-65 to +150	°C
T_A	Operating temperature range	0 to +70	°C
T_{SOLD}	Lead temperature (soldering, 10sec)	300	°C
I_{SC}	Output short-circuit duration	Indefinite	

NOTES:

1. Derate above 25°C at the following rates:
   N package at 12mW/°C
2. For supply voltages less than ±15V, the absolute maximum input voltage is equal to the supply voltage.

DC ELECTRICAL CHARACTERISTICS

T_A=25°C, V_{CC} = ±15V unless otherwise specified.

SYMBOL	PARAMETER	TEST CONDITIONS	µA747C			UNIT
			Min	Typ	Max	
V_{OS}	Offset voltage	R_S≤10kΩ		2.0	6.0	mV
		R_S≤10kΩ, over temp.		3.0	7.5	mV
$\Delta V_{OS}/\Delta T$				10		µV/°C
I_{OS}	Offset current			20	200	nA
		Over temperature		7.0	300	nA
$\Delta I_{OS}/\Delta T$				200		pA/°C
I_{BIAS}	Input current			80	500	nA
		Over temperature		30	800	nA
$\Delta I_B/\Delta T$				1		nA/°C
V_{OUT}	Output voltage swing	R_L≥2kΩ, over temp.	±10	±13		V
		R_L≥10kΩ, over temp.	±12	±14		V
I_{CC}	Supply current each side			1.7	2.8	mA
		Over temperature		2.0	3.3	mA
P_d	Power consumption			50	85	mW
		Over temperature		60	100	mW
C_{IN}	Input capacitance			1.4		pF
	Offset voltage adjustment range			±15		mV
R_{OUT}	Output resistance			75		Ω
	Channel separation			120		dB
PSRR	Supply voltage rejection ratio	R_S≤10kΩ, over temp.		30	150	µV/V
A_{VOL}	Large-signal voltage gain (DC)	R_L≥2kΩ, V_{OUT}=±10V	25,000			V/V
		Over temperature	15,000			V/V
CMRR	Common-mode rejection ratio	R_S≤10kΩ, V_{CM}=±12V Over temperature	70			dB

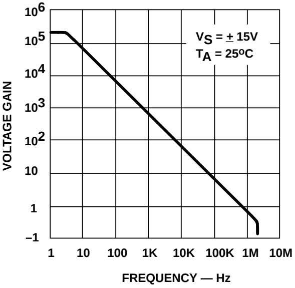
Dual operational amplifier µA747C

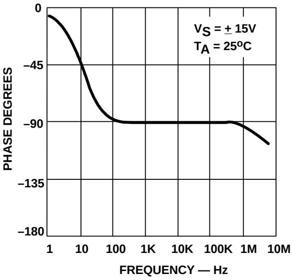
AC ELECTRICAL CHARACTERISTICS

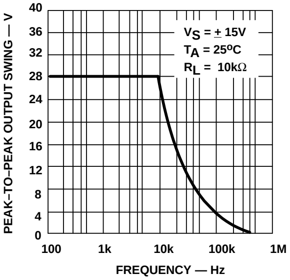
T_A=25°C, V_S = ±15V unless otherwise specified.

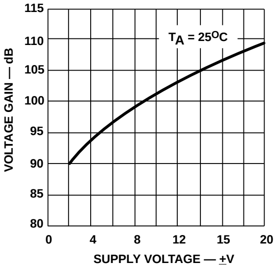
SYMBOL	PARAMETER	TEST CONDITIONS	µA747C			UNIT
			Min	Typ	Max	
	Transient response	V_{IN}=20mV, R_L=2kΩ, C_L<100pF				
t_R	Rise time	Unity gain C_L≤100pF		0.3		µs
	Overshoot	Unity gain C_L≤100pF		5.0		%
SR	Slew rate	R_L>2kΩ		0.5		V/µs

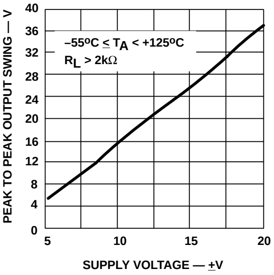
TYPICAL PERFORMANCE CHARACTERISTICS

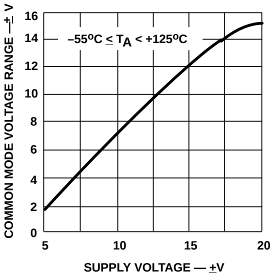

Open–Looped Voltage Gain as a Function of Frequency

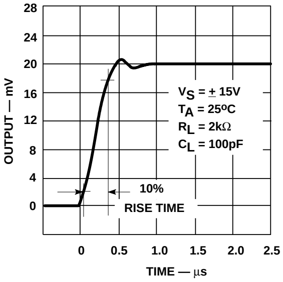

Open–Looped Voltage Response as a Function of Frequency

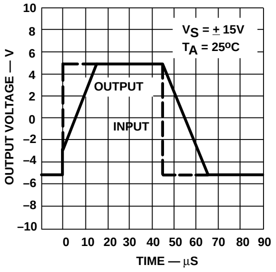

Output Voltage Swing as a Function of Frequency

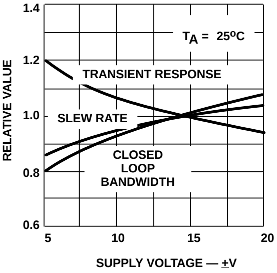

Open–Loop Voltage Gain as a Function of Supply Voltage


Output Voltage Swing as a Function of Supply Voltage


Input Common–Mode Voltage Range as a Function of Supply voltage


Transient Response

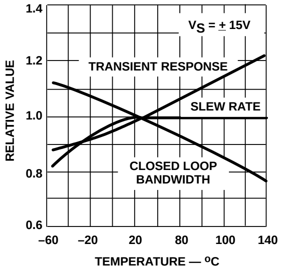

Voltage-follower Large-Signal Pulse Response


Frequency Characteristics as a Function of Supply Voltage

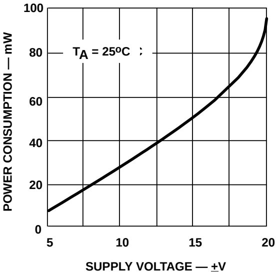
Dual operational amplifier μA747C

TYPICAL PERFORMANCE CHARACTERISTICS (Continued)

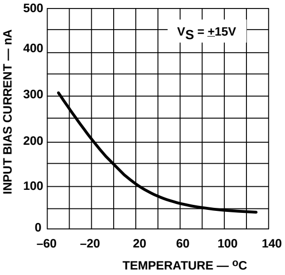
Frequency Characteristics as a Function of Ambient Temperature



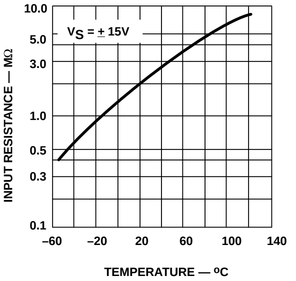
Power Consumption as a Function of Supply Voltage



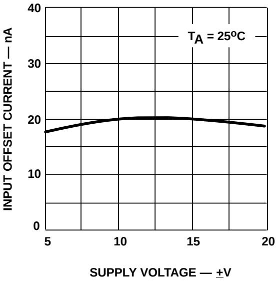
Input Bias Current as a Function of Ambient Temperature



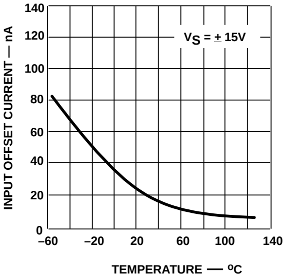
Input Resistance as a Function of Ambient Temperature



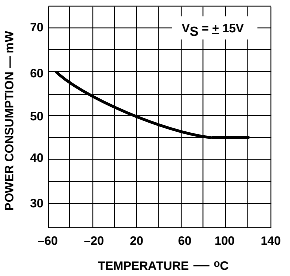
Input Offset Current as a Function of Supply Voltage



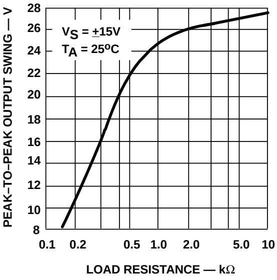
Input Offset Current as a Function of Ambient Temperature



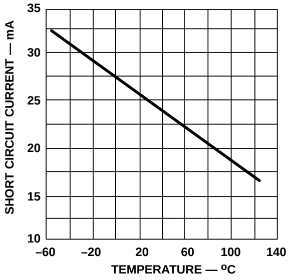
Power Consumption as a Function of Ambient Temperature



Output Voltage Swing as a Function of Load Resistance

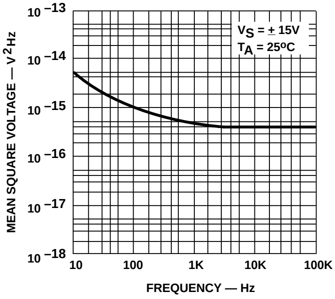


Output Short–Circuit Current as a Function of Ambient Temperature

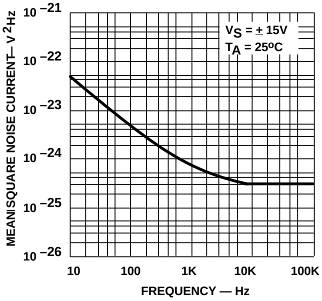


TYPICAL PERFORMANCE CHARACTERISTICS (Continued)

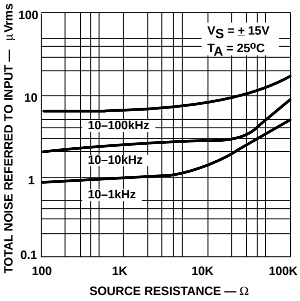
Input Noise Voltage as a Function of Frequency



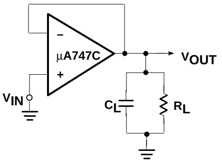
Input Noise Current as a Function of Frequency



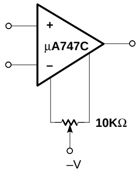
Broadband Noise for Various Bandwidths



TEST CIRCUITS

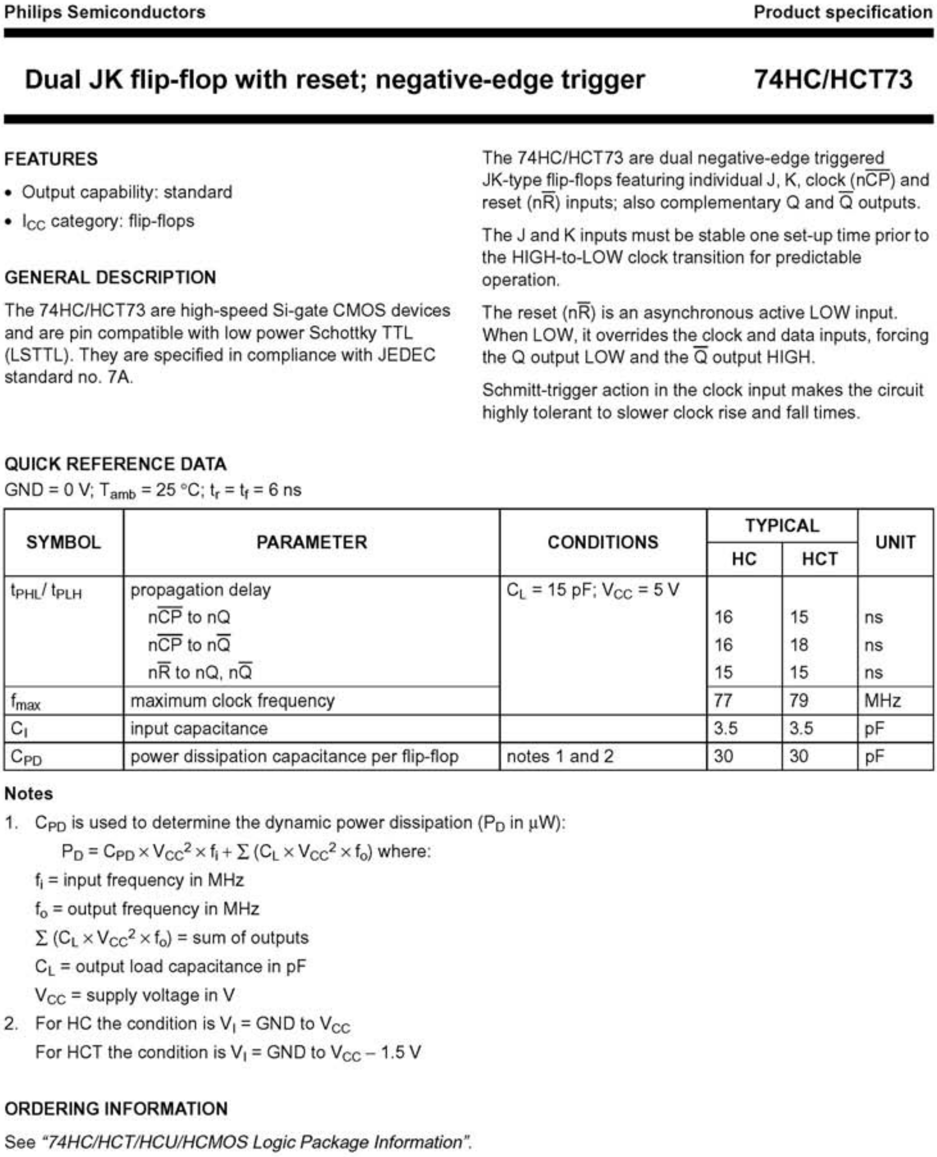


Transient Response Test Circuit



Voltage Offset Null Circuit

A.2.2 Specifications of the 74HCT73 (a digital circuit)



Philips Semiconductors **Product specification**

Dual JK flip-flop with reset; negative-edge trigger 74HC/HCT73

FEATURES

- Output capability: standard
- I_{CC} category: flip-flops

GENERAL DESCRIPTION

The 74HC/HCT73 are high-speed Si-gate CMOS devices and are pin compatible with low power Schottky TTL (LSTTL). They are specified in compliance with JEDEC standard no. 7A.

The 74HC/HCT73 are dual negative-edge triggered JK-type flip-flops featuring individual J, K, clock ($n\overline{CP}$) and reset ($n\overline{R}$) inputs; also complementary Q and $\overline{Q}$ outputs.

The J and K inputs must be stable one set-up time prior to the HIGH-to-LOW clock transition for predictable operation.

The reset ($n\overline{R}$) is an asynchronous active LOW input. When LOW, it overrides the clock and data inputs, forcing the Q output LOW and the $\overline{Q}$ output HIGH.

Schmitt-trigger action in the clock input makes the circuit highly tolerant to slower clock rise and fall times.

QUICK REFERENCE DATA

GND = 0 V; T_{amb} = 25 °C; $t_r = t_f$ = 6 ns

SYMBOL	PARAMETER	CONDITIONS	TYPICAL		UNIT
			HC	HCT	
t_{PHL}/t_{PLH}	propagation delay	C_L = 15 pF; V_{CC} = 5 V			
	$n\overline{CP}$ to nQ		16	15	ns
	$n\overline{CP}$ to $n\overline{Q}$		16	18	ns
	$n\overline{R}$ to nQ, $n\overline{Q}$		15	15	ns
f_{max}	maximum clock frequency		77	79	MHz
C_I	input capacitance		3.5	3.5	pF
C_{PD}	power dissipation capacitance per flip-flop	notes 1 and 2	30	30	pF

Notes

1. C_{PD} is used to determine the dynamic power dissipation (P_D in μW):

   $P_D = C_{PD} \times V_{CC}^2 \times f_i + \sum (C_L \times V_{CC}^2 \times f_o)$ where:

   f_i = input frequency in MHz

   f_o = output frequency in MHz

   $\sum (C_L \times V_{CC}^2 \times f_o)$ = sum of outputs

   C_L = output load capacitance in pF

   V_{CC} = supply voltage in V

2. For HC the condition is V_I = GND to V_{CC}

   For HCT the condition is V_I = GND to V_{CC} – 1.5 V

ORDERING INFORMATION

See *"74HC/HCT/HCU/HCMOS Logic Package Information"*.

Dual JK flip-flop with reset; negative-edge trigger 74HC/HCT73

PIN DESCRIPTION

PIN NO.	SYMBOL	NAME AND FUNCTION
1, 5	$1\overline{CP}$, $2\overline{CP}$	clock input (HIGH-to-LOW, edge-triggered)
2, 6	$1\overline{R}$, $2\overline{R}$	asynchronous reset inputs (active LOW)
4	V_{CC}	positive supply voltage
11	GND	ground (0 V)
12, 9	1Q, 2Q	true flip-flop outputs
13, 8	$1\overline{Q}$, $2\overline{Q}$	complement flip-flop outputs
14, 7, 3, 10	1J, 2J, 1K, 2K	synchronous inputs; flip-flops 1 and 2

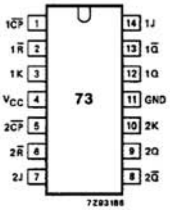

$1\overline{CP}$ 1
$1\overline{R}$ 2
1K 3
V_{CC} 4
$2\overline{CP}$ 5
$2\overline{R}$ 6
2J 7
73
14 1J
13 $1\overline{Q}$
12 1Q
11 GND
10 2K
9 2Q
8 $2\overline{Q}$
7Z93186

Fig.1 Pin configuration.

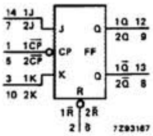

14 1J
7 2J
1 $1\overline{CP}$
5 $2\overline{CP}$
3 1K
10 2K
J
CP
FF
K
Q
$\overline{Q}$
R
1Q 12
2Q 9
$1\overline{Q}$ 13
$2\overline{Q}$ 8
$1\overline{R}$ $2\overline{R}$
2 6
7Z93187

Fig.2 Logic symbol.

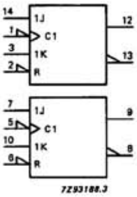

14
1
3
2
1J
C1
1K
R
12
13
7
5
10
6
1J
C1
1K
R
9
8
7Z93188.3

Fig.3 IEC logic symbol.

Philips Semiconductors Product specification

Dual JK flip-flop with reset; negative-edge trigger 74HC/HCT73

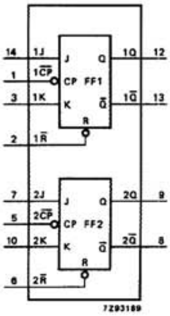

14
1J
J
Q
1Q
12
1
1$\overline{CP}$
CP FF1
3
1K
K
$\overline{Q}$
1$\overline{Q}$
13
R
2
1$\overline{R}$
7
2J
J
Q
2Q
9
5
2$\overline{CP}$
CP FF2
10
2K
K
$\overline{Q}$
2$\overline{Q}$
8
R
6
2$\overline{R}$
7Z93189

Fig.4 Functional diagram.

FUNCTION TABLE

OPERATING MODE	INPUTS				OUTPUTS	
	n$\overline{R}$	n$\overline{CP}$	J	K	Q	$\overline{Q}$
asynchronous reset	L	X	X	X	L	H
toggle	H	↓	h	h	$\overline{q}$	q
load "0" (reset)	H	↓	l	h	L	H
load "1" (set)	H	↓	h	l	H	L
hold "no change"	H	↓	l	l	q	$\overline{q}$

Notes

1. H = HIGH voltage level
   h = HIGH voltage level one set-up time prior to the HIGH-to-LOW CP transition
   L = LOW voltage level
   l = LOW voltage level one set-up time prior to the HIGH-to-LOW CP transition
   q = lower case letters indicate the state of the referenced output one set-up time prior to the HIGH-to-LOW CP transition
   X = don't care
   ↓ = HIGH-to-LOW CP transition

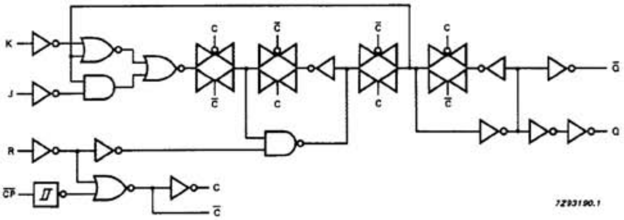

K
J
R
$\overline{CP}$
c
$\overline{c}$
$\overline{Q}$
Q
7Z93190.1

Fig.5 Logic diagram (one flip-flop).

Philips Semiconductors Product specification

Dual JK flip-flop with reset; negative-edge trigger 74HC/HCT73

DC CHARACTERISTICS FOR 74HC

For the DC characteristics see *"74HC/HCT/HCU/HCMOS Logic Family Specifications"*.

Output capability: standard
I_{CC} category: flip-flops

AC CHARACTERISTICS FOR 74HC

GND = 0 V; $t_r = t_f = 6$ ns; $C_L = 50$ pF

SYMBOL	PARAMETER	T_{amb} (°C)							UNIT	TEST CONDITIONS	
		74HC								V_{CC} (V)	WAVEFORMS
		+25			−40 to +85		−40 to +125				
		min.	typ.	max.	min.	max.	min.	max.			
t_{PHL}/ t_{PLH}	propagation delay $n\overline{CP}$ to nQ		52 19 15	160 32 27		200 40 34		240 48 41	ns	2.0 4.5 6.0	Fig.6
t_{PHL}/ t_{PLH}	propagation delay $n\overline{CP}$ to $n\overline{Q}$		52 19 15	160 32 27		200 40 34		240 48 41	ns	2.0 4.5 6.0	Fig.6
t_{PHL}/ t_{PLH}	propagation delay $n\overline{R}$ to nQ, $n\overline{Q}$		50 18 14	145 29 25		180 36 31		220 44 38	ns	2.0 4.5 6.0	Fig.7
t_{THL}/ t_{TLH}	output transition time		19 7 6	75 15 13		95 19 16		110 22 19	ns	2.0 4.5 6.0	Fig.6
t_W	clock pulse width HIGH or LOW	80 16 14	22 8 6		100 20 17		120 24 20		ns	2.0 4.5 6.0	Fig.6
t_W	reset pulse width HIGH or LOW	80 16 14	22 8 6		100 20 17		120 24 20		ns	2.0 4.5 6.0	Fig.7
t_{rem}	removal time $n\overline{R}$ to $n\overline{CP}$	80 16 14	22 8 6		100 20 17		120 24 20		ns	2.0 4.5 6.0	Fig.7
t_{su}	set-up time nJ, nK to $n\overline{CP}$	80 16 14	22 8 6		100 20 17		120 24 20		ns	2.0 4.5 6.0	Fig.6
t_h	hold time nJ, nK to $n\overline{CP}$	3 3 3	−8 −3 −2		3 3 3		3 3 3		ns	2.0 4.5 6.0	Fig.6
f_{max}	maximum clock pulse frequency	6.0 30 35	23 70 83		4.8 24 28		4.0 20 24		MHz	2.0 4.5 6.0	Fig.6

Philips Semiconductors | Product specification

Dual JK flip-flop with reset; negative-edge trigger | 74HC/HCT73

DC CHARACTERISTICS FOR 74HCT

For the DC characteristics see *"74HC/HCT/HCU/HCMOS Logic Family Specifications"*.

Output capability: standard
I_{CC} category: flip-flops

Note to HCT types

The value of additional quiescent supply current (ΔI_{CC}) for a unit load of 1 is given in the family specifications. To determine ΔI_{CC} per input, multiply this value by the unit load coefficient shown in the table below.

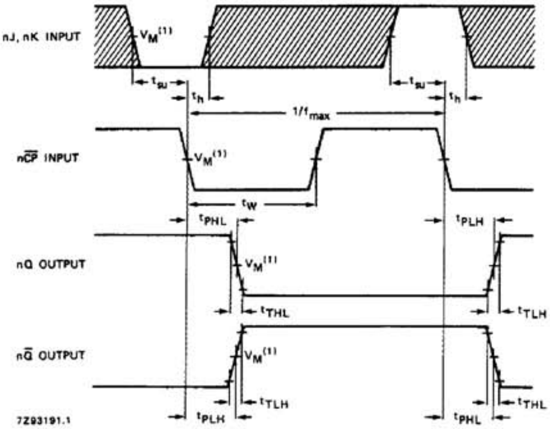
INPUT	UNIT LOAD COEFFICIENT
nK	0.60
$n\overline{R}$	0.65
$n\overline{CP}$, nJ	1.00

AC CHARACTERISTICS FOR 74HCT

GND = 0 V; $t_r = t_f = 6$ ns; $C_L = 50$ pF

SYMBOL	PARAMETER	T_{amb} (°C)							UNIT	TEST CONDITIONS	
		74 HCT								V_{CC} (V)	WAVEFORMS
		+25			−40 to +85		−40 to +125				
		min.	typ.	max.	min.	max.	min.	max.			
t_{PHL}/t_{PLH}	propagation delay $n\overline{CP}$ to nQ		18	38		48		57	ns	4.5	Fig.6
t_{PHL}/t_{PLH}	propagation delay $n\overline{CP}$ to $n\overline{Q}$		21	36		45		54	ns	4.5	Fig.6
t_{PHL}/t_{PLH}	propagation delay $n\overline{R}$ to nQ, $n\overline{Q}$		20	34		43		51	ns	4.5	Fig.7
t_{THL}/t_{TLH}	output transition time		7	15		19		22	ns	4.5	Fig.6
t_W	clock pulse width HIGH or LOW	16	8		20		24		ns	4.5	Fig.6
t_W	reset pulse width HIGH or LOW	18	9		23		27		ns	4.5	Fig.7
t_{rem}	removal time $n\overline{R}$ to $n\overline{CP}$	14	8		18		21		ns	4.5	Fig.7
t_{su}	set-up time nJ, nK to $n\overline{CP}$	12	6		15		18		ns	4.5	Fig.6
t_h	hold time nJ, nK to $n\overline{CP}$	3	−2		3		3		ns	4.5	Fig.6
f_{max}	maximum clock pulse frequency	30	72		24		20		MHz	4.5	Fig.6

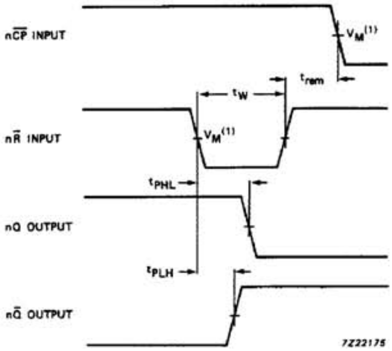
AC WAVEFORMS


nJ, nK INPUT
$V_M^{(1)}$
t_{su}
t_h
t_{su}
t_h
$1/f_{max}$
$n\overline{CP}$ INPUT
$V_M^{(1)}$
t_W
t_{PHL}
t_{PLH}
nQ OUTPUT
$V_M^{(1)}$
t_{THL}
t_{TLH}
$n\overline{Q}$ OUTPUT
$V_M^{(1)}$
t_{TLH}
t_{THL}
t_{PLH}
t_{PHL}
7Z93191.1

The shaded areas indicate when the input is permitted to change for predictable output performance.

(1) HC : V_M = 50%; V_I = GND to V_{CC}.
HCT: V_M = 1.3 V; V_I = GND to 3 V.

Fig.6 Waveforms showing the clock ($n\overline{CP}$) to output (nQ, $n\overline{Q}$) propagation delays, the clock pulse width, the J and K to $n\overline{CP}$ set-up and hold times, the output transition times and the maximum clock pulse frequency.


$n\overline{CP}$ INPUT
$V_M^{(1)}$
t_{rem}
t_W
$n\overline{R}$ INPUT
$V_M^{(1)}$
t_{PHL}
nQ OUTPUT
t_{PLH}
$n\overline{Q}$ OUTPUT
7Z22175

(1) HC : V_M = 50%; V_I = GND to V_{CC}.
HCT: V_M = 1.3 V; V_I = GND to 3 V.

Fig.7 Waveforms showing the reset ($n\overline{R}$) input to output (nQ, $n\overline{Q}$) propagation delays and the reset pulse width and the $n\overline{R}$ to $n\overline{CP}$ removal time.

PACKAGE OUTLINES

See *"74HC/HCT/HCU/HCMOS Logic Package Outlines"*.

Answers to exercises

The answers to the exercises can be downloaded from http://www.vssd.nl/hlf/e008.htm or requested by sending an e-mail tot hlf@vssd.nl.

Index

#

−20dB/decade, 82
−3dB frequency, 81
−6dB/octave, 82

A

absolute value, 50
acceleration, 105
accelerometer
  piezoelectric, 110
acceptor, 131
accuracy, 20
acquisition, 14
actuator, 16
ADC, 289
  compensating, 296
  direct, 301
  dual-slope, 303
  integrating, 302
  parallel, 296
  successive approximation, 296
AD-conversion, 15
AD-converters. *See* ADC
adder, 321
  current, 185
  full, 322
  half, 321
  voltage, 185
additive errors, 72
admittance, 52
alias, 26
aliasing error, 26
alpha-numeric display, 136
alternate mode, 340
AM, 271
  frequency spectrum, 274, 281
  time signal, 273
Ampère's law, 98
amplification, 16
amplifier
  buffer, 186
  chopper, 284
  current, 149
  differential, 162, 182, 186
  electrometer, 285
  indirect, 285
  instrumentation, 188
  inverting, 185
  lock-in, 284
  non-inverting, 186
  operational, 182
  varactor, 285
amplitude characteristic, 80
amplitude control, 257
amplitude modulation, 271
amplitude stabilization, 260
amplitude transfer, 54
analog multiplier, 225
analogue multiplexer, 243
analogue multiplier, 275
analogue-to-digital converter, 289
AND, 309, 314, 321
angular velocity, 44
anti-log converter, 225
aperture delay time, 247
aperture jitter, 247
aperture uncertainty, 247
approximation
  asymptotic, 82
  Bode plots, 80
  piecewise linear, 229
  polar plots, 87
arbitrary power function, 229
argument, 50, 53, 80
asymptotic approximation, 82
attenuator
  probe, 342
auto-polarity, 337
auto-ranging, 337
auto-scaling, 337
average
  time, 32
  weighted, 29

B

band-pass filter, 122, 124
bandwidth, 23, 20, 122
  relative, 84
  signal, 21
  unity gain, 195
base, 148
base resistance, 152
BCD, 291
bel, 71
Bessel filter, 126
bias
  base-current, 154
  base-voltage, 157
bias current, 182, 183, 191
  compensation, 192
bias point, 152
biasing, 151, 156, 174
binary code, 289
binary counter, 325
binary word, 290
Biot and Savart law, 98
bipolar transistor, 148, 151
bit, 289
blanking, 338
Bode plot, 80, 345, 347
Bode relation, 127
Boltzmann's constant, 133
Boolean algebra, 308
branch, 38
breakthrough voltage, 96
bridge
  Graetz, 144
  Wheatstone, 279, 366
bridged-T filter, 123
buffer amplifier, 186, 194
bus, 348
bus protocol, 348

bus structure, 348
Butterworth filter, 125
byte, 290

C

capacitance, 95, 106
  generalized, 43
  heat, 44
  impedance, 52
  mechanical, 44
  stray, 363
capacitive sensors, 106
capacitor, 95
  adjustable, 96
  ceramic, 97
  coupling, 155
  electrolytic, 96
  mica, 97
  trimming, 96
  variable, 96
  wet aluminum, 97
carbon film resistor, 94
carrier, 271, 273
carry bit, 321
cascaded DAC, 301
cathode, 337
cathode ray tube, 337
channel, 167, 173
characteristic impedance, 69, 344
Chebychev filter, 126
chopped mode, 340
chopper amplifier, 284
clamp circuit, 140
clipper, 137
clipping, 21
clock, 317
CMOS, 316
CMRR, 24, 163, 187
cold junction, 108
collector, 148
combinatory circuit, 313
command-driven, 354
common mode, 23, 163, 187
common mode rejection ratio, 24, 163, 187
common-emitter, 155
comparator, 215, 301
compensating ADC, 296
compensation, 366
  frequency compensation, 195
complementary MOS, 316
complex plane, 50, 52, 85
complex variables, 50
component
  active, 37, 93
  electronic, 93
  passive, 37, 93
conductivity, 93, 131
  of LDR, 103
  values, 94
conductors, 94
conjugate, 85
conversion
  analog-to-digital, 15
  digital-to-analog, 16
converter
  anti-log, 225
  current-to-voltage, 184
  exponential, 225
  logarithmic, 222
  log-ratio, 224
  voltage-to-current, 152, 174
  voltage-to-frequency, 267
convolution, 25
correction, 367
counter, 323, 344
  asynchronous, 323
  binary, 325
  decimal, 328
  digital, 323
  ripple, 323
coupling capacitor, 155, 157, 162
covalent bond, 130
cross talk, 72
crosstalk, 248
crystal oscillator, 343
Curie temperature, 110
current density, 93
current gain, 149, 154
current matching, 69
current source
  voltage controlled, 149
current switch, 235
current-to-voltage converter, 184

D

DA conversion, 16
DAC
  cascaded, 301
  multiplying, 306
  serial, 300
DA-converter. *See* DAC
damping factor, 256
damping ratio, 84
dark current, 135
dark resistance, 103
data acquisition, 14
data distribution, 15
data driven, 354
data processing, 14
data ready, 299
dB, 71
DC signal source, 343
De Morgan's theorem, 312
dead zone, 21
decade, 82
decibel, 71, 81
deflection plates, 337
demodulation, 16, 271
demodulator
  FM, 282
demultiplexer, 244
depletion layer, 131
detection, 279. *See* demodulation
  synchronous, 279
detector
  peak, 139
  peak-to-peak, 143
dielectric, 95
dielectric constant, 95, 106
  table, 96
dielectric loss, 96
differential amplifier, 23, 163, 182, 186, 366
differential emitter resistance, 151
erential equation, 39, 50, 54
  constant coefficients, 39
  linear, 39
  n-th order, 57
  diffordinary, 40
differential mode, 23, 187
differential non-linearity, 295
differential resistance, 134
differentiating network, 120
differentiator, 256
digital multiplexer, 320
digital signal, 289
digital-to-analogue converter, 289
diode, 133
  light emitting, 136
  light sensitive, 135
direct ADC, 301
displacement, 105
distribution function, 27, 359
  Gaussian, 28, 359
  normal, 28, 30, 31, 359
divider, 228
donor, 131, 172
double-sided rectifier, 144, 221, 279
down-counter, 325
drain, 167, 173
drift, 22, 14, 367
droop, 247
dual, 314
dual integrator loop, 260
dual-slope ADC, 303
duty cycle, 263

E

E-12, 95
E-48, 95
Early effect, 150, 171
eddy current, 106
effector, 16
electrolytic capacitor, 96
electronic switch, 235
element
  active, 63
  dissipating, 45

  electric network, 37
  model, 63
  passive, 63
emissivity diagram, 136
emitter, 148
emitter follower, 160
emitter resistance
  differential, 151
enable, 243, 320, 325
envelope, 273
error
  absolute, 361
  additive, 72
  aliasing, 26
  destructive, 72
  equivalent, 73
  multiplicative, 72
  random, 358
  relative, 68, 361
  systematic, 358
  transient, 248
error propagation, 361
estimation, 359
Euler, 22
excitator, 16
exclusive OR, *see* EXOR
EXOR, 310, 314, 321
expectancy, 29
exponential converter, 225
external frequency compensation, 195

F

factorization, 83
failure rate, 24
fall time, 236
farad, 95
Faraday's law of induction, 97
FET, 167
  n-channel, 167
  p-channel, 167
field-effect transistor, 167
filter
  active, 115
  approximation, 125
  band-pass, 122
  Bessel, 126
  bridged-T, 123
  Butterworth, 125
  Chebychev, 126
  high-pass, 119
  inductorless, 116
  low-pass, 116
  notch, 123
  passive, 115
  predetection, 284
  RC-, 116
filtering, 16, 367
Firewire, 349
first order moment, 29
flipflop, 316
  JK, 318
  master-slave, 318
  SR, 316
flux
  magnetic, 97
FM, 271
FM-demodulator, 282
force sensitive resistors, 103
forward voltage, 133
Fourier coefficients, 18
Fourier integral, 25, 59
Fourier series, 17, 276
  complex, 22
Fourier transform, 26, 32, 54
  discrete, 19, 23
free charge carriers, 130
free electrons, 130
frequency
  –3dB, 81
frequency band, 23
frequency conversion, 271
frequency divider, 318
frequency meter, 344
frequency modulation, 271
frequency multiplexing, 274
frequency spectrum
  AM, 274, 281
  continuous, 20
full-adder, 322
fundamental frequency, 17

G

gain
  current, 149
gallium arsenide, 130, 136
gate, 168, 173
  logic, 313
gauge factor, 104
Gaussian distribution, 359
GB-product, 195
generalized network, 42
generator
  function, 343
  noise, 344
  pulse, 265, 343
  ramp, 345
  ramp voltage, 263
  signal, 343
  sine wave, 255, 343
  square wave, 265, 343
  sweep, 266, 343
  triangle, 261
  voltage, 261
glass fiber, 136
GPIB, 348
Graetz bridge, 144, 221
ground
  circuit, 364
  instrument, 364
  safety, 364
ground loop, 364
guard connection, 368
guarding
  active, 368

H

half-adder, 321
handshake, 350
harmonic components, 18
harmonic oscillator, 255, 258
heat capacitance, 44
hexadecimal code, 291
high-pass filter, 119
histogram, 360
hold capacitor, 247
hold range, 283
hole, 130
hot junction, 108
HP-IB, 348
humidity sensor, 351
hysteresis, 99, 217, 265

I

IEC, 95, 348
IEC 625, 349
IEC bus, 349, 351
IEEE 1394, 349
IEEE 488, 348
imaginary unit, 50
impedance, 52, 58
  characteristic, 69
  input, 64
  output, 64
  source, 63
impedance analyzer, 347, 352
inaccuracy, 20
  absolute, 20, 361
  relative, 20, 361
indirect DC amplifier, 285
inductance
  mutual, 98
  self-, 98
induction, 97
inductive sensors, 104
inductor, 97
input impedance, 64, 66
input offset, 22
instrument
  embedded, 336
  virtual, 348
instrumentation amplifier, 188
integrating ADC, 302
integrating network, 117
interference, 72, 359, 362
International Electrotechnical Commission, 313, 348
inversion, 310
inversion layer, 173

inverter, 314
inverting amplifier, 185
isolation, 365
isolator, 94

J

JFET, 167
jitter, 246
JK-flipflop, 318
junction
  cold, 108
  hot, 108
  pn, 132, 148

K

Kirchhoff, 38
Kirchhoff's rules, 38

L

ladder network, 293
Laplace operator, 55, 59
Laplace transform, 54, 56
law of inertia, 44
LDR, 102
leakage current, 133, 150
least significant bit, 290
LED, 136
Lenz's law, 98
level sensors, 106
level shifter, 181
light sensitive resistors, 102
light-emitting diode, 136
limiter, 137, 219
Linear displacement sensor, 104
linear variable differential transformer, 105
listen, 349
lock-in amplifier, 284
lock-in range, 283
logarithmic voltage converter, 222
logic gate, 313
log-ratio converter, 224
long-tailed pair, 163
loop, 38
loss angle, 96, 97
low-pass filter, 116
LSB, 290
lumped element, 43
LVDT, 105

M

magnetic flux, 97
magnetic induction, 97, 363, 366
master-slave flipflop, 318
matching, 68, 344
  characteristic, 70
  current, 69
  power, 70
  voltage, 68
mean absolute value, 16
mean signal power, 16
mean value, 16
mean-time-to-failure, 24
measurement error, 358
measurement uncertainty, 358
mechanical damper, 44
metal film resistor, 94
metal wire resistor, 94
mobility, 100
model
  current source, 64
  voltage source, 64
modulation, 16, 271, 367
  amplitude, 271
  frequency, 271
  phase, 271
  pulse height, 271
  pulse width, 271
  single sideband, 275
  suppressed carrier, 275
modulation depth, 273
modulator, 100
  switch, 276
  Wheatstone bridge, 279
modulus, 50, 53, 80
moment of torsion, 44
monotony, 295
MOSFET, 171
  normally on, 173
  normally-off, 173
most significant bit, 290
MSB, 290
MTTF, 24
multi-channel, 16
multi-channel oscilloscope, 340
multimeter, 337
multiplexer
  analogue, 243
  differential, 243
  digital, 320
  time, 243
multiplexing, 17
  digital, 16
  frequency, 16, 18, 274
  time, 16
multiplier
  analogue, 225, 275
multi-turn potentiometer, 104
mutual inductance, 98, 104

N

NAND, 314
negation, 310
network
  differentiating, 120
  generalized, 42
  integrating, 117
  one-port, 40
  three terminal, 40
  two-port, 40
  two-terminal, 40
network analyzer, 345
Newton, 44
node, 38
noise, 23, 22, 72, 75, 173, 359, 362
  thermal, 75
  white, 22
noise generator, 344
noise power, 75
non-inverting amplifier, 186
non-linearity, 21
  differential, 295
normal distribution, 359
Norton's, 64
NOT, 310, 314
notch filter, 123
NTC, 102

O

object oriented programming, 354
octal code, 291
octave, 82
off-resistance, 237
offset, 22, 72
  compensation, 192
  input, 73
  output, 73
offset current, 193
offset voltage, 182, 183, 190
ohm, 94
Ohm's law, 44, 94
on-resistance, 237
open loop gain, 194
open voltage, 63
operating range, 19
operational amplifier, 182
OR, 309, 314
oscillation condition, 258
oscillator, 255
  crystal, 343
  harmonic, 258
  phase shift, 259
  sine wave, 255
  two-integrator, 260
  voltage controlled, 266, 343
  Wien, 258
oscilloscope
  digital, 340
  dual beam, 340
  multi-channel, 340
  probe, 341
OSI/ISO, 348
output impedance, 64, 66
overshoot, 84

P

parallel ADC, 296
parallel converter, 289
parallel word, 292
peak detector, 139, 144, 279
peak value, 16
peak wavelength, 136
peak-to-peak detector, 143
peak-to-peak value, 16
Peltier coefficient, 109
Peltier effect, 109
permeability, 99
  of vacuum, 99
  relative, 99
permittivity, 95
  absolute, 95
  of vacuum, 95
  table, 96
phase, 80
phase characteristic, 80
phase modulation, 271
phase transfer, 54
phase-locked loop, *see* PLL
phase-shift oscillator, 259
photodiode, 134
piecewise linear approximation, 229
piezoelectric accelerometer, 110
piezoelectric sensors, 110
piezoelectricity, 110
piezoresistive effect, 103
pinch-off, 170
pinch-off voltage, 168
pipe-lining, 301
platinum resistance thermometer, 101
PLL, 282, 283, 284
pn-diode, 130
pn-junction, 132, 148
polar coordinates, 50
polar plot, 85
pole, 58
poling, 110
port
  input, 40
  output, 40
potentiometer, 104
power density
  spectral, 75
power density spectrum, 32, 75
power spectrum, 20
power supply, 343
power transfer, 71
predetection filter, 284
probability, 27
probability density, 28
probability density function, 360
probes, 341
programmable chip, 333
PROM, 352
Pt-100, 101
PTC, 102
pulse generator, 265
pulse height modulation, 271
pulse width modulation, 271

Q

quad, 314
quality factor, 84
quantization, 15
quantum efficiency, 136

R

ramp generator, 263
random error, 358
rectifier
  double-sided, 144, 221
  single-sided, 221, 280
reed switch, 239
relative bandwidth, 84
relative permeability, 99
reliability, 24
resistance, 94
  base, 152
  differential, 134
  generalized, 44
  hydraulic, 44
  input, 158, 161
  output, 158, 161
  thermal, 44
resistance thermometers, 101
resistive displacement sensor, 104
resistive sensors, 100
resistivity, 93, 100
resistor, 93
  adjustable, 94
  fixed, 94
  force sensitive, 103
  light sensitive, 102
  properties, 94
  variable, 94
resolution, 19, 104
resonance frequency, 343
reverse current, 133
reverse voltage, 133
ripple, 140, 144
ripple counter, 323
rise time, 236
rms, 73
root-mean-square value, 16
ruler, 104

S

sample-hold, 245, 247, 292
sampling, 15, 292
sampling theorem, 26
SAR, 298
saturation line, 241
Schmitt-trigger, 218, 261, 263, 344
second order moment, 29
Seebeck coefficient, 108, 110
Seebeck effect, 107
Seebeck voltage, 107
self-inductance, 98, 99, 104
  generalized, 44
  impedance, 52
semiconductor, 94
  intrinsic, 100
sensitivity, 20
sensor, 15, 100
  capacitive, 106
  concentration, 106
  displacement, 104, 106
  eddy current, 106
  inductive, 104, 105
  level, 106
  piezoelectric, 110
  resistive, 100
  thermoelectric, 107, 108
sequential circuit, 313, 316
serial DAC, 300
serial word, 292
series switch, 237, 241
series-shunt switch, 238
settling time, 236
seven-segment code, 329
seven-segment decoder, 329
Shannon, 26
shielding, 72, 365, 369
shift register, 325
short-circuit current, 63
shunt switch, 238, 240, 241
side band, 274
siemens, 94
signal
  AC, 14
  analog, 15
  aperiodic, 22
  binary, 15
  continuous, 15
  DC, 14
  deterministic, 14
  digital, 15, 289
  discrete, 15
  dynamic, 14
  interference, 72
  periodic, 14
  quantized, 15
  quasi stochastic, 344
  quasi-static, 14
  sampled, 15, 26
  static, 14
  stochastic, 14, 27
  transients, 15
signal generator, 343
silicon
  intrinsic, 130
  n-type, 131

  p-type, 131
sine wave generator, 343
sine wave oscillator, 255
single sideband modulation, 275
single-sided rectifier, 221, 280
slew rate, 21
source, 167, 173
source follower, 176
source impedance, 63
source resistance, 73
spectral response, 135
spectrum, 17
  complex, 23
  line, 20
spectrum analyzer, 345
square root transfer, 228
square summing rule, 74
square wave generator, 265
SR-flipflop, 316
standard deviation, 29, 32, 360, 361
step response, 57, 117, 119
stiffness of rotation, 44
strain gauge, 103, 366
strain gauge factor, 104
stray capacitance, 363
stray fields, 364
strobe, 217
substrate, 173
successive approximation ADC, 296
successive approximation register, 298
supermalloy, 99
superposition, 187, 191
suppressed carrier, 275, 277
sweep generator, 266, 343
switch
  bipolar transistor, 240
  current, 235
  diode bridge, 240
  electronic, 235
  floating, 235
  junction FET, 241
  MOSFET, 242
  photo resistor, 239
  pn-diode, 239
  reed, 239
  series, 237
  series-shunt, 238
  shunt, 238
  voltage, 235
switch modulator, 276
synchronous detection, 279, 284, 345
systematic error, 358

T

T filter, 123
tachometer, 106
talk, 349
temperature coefficient, 22, 94
temperature sensor, 100, 108
T-equivalent circuit, 152
tesla, 97
thermistor, 101
thermocouple, 108
  sensitivity, 109
thermoelectric sensors, 107
thermopile, 109
Thévenin, 63
Thévenin voltage, 64
Thomson coefficient, 110
Thomson effect, 109, 110
threshold voltage, 133
thyristor, 242
time base, 337
time constant, 80, 117
time meter, 344
time multiplexer, 243
toggle, 318, 323
track-hold, 245
transconductance, 40, 151, 171
transducer, 100
  input, 15
  output, 16
transfer
  logarithmic, 71
transfer function, 58
  complex, 52
transformer, 97, 100
  LVDT, 105
transient error, 248
transistor
  bipolar, 148
  field effect, 167
  model, 152
  npn, 148
  pnp, 148, 149
  super-β, 149
Transistor-Transistor Logic, 314
triac, 242
triangle generator, 261
trigger
  automatic, 338
  delayed, 338
  external, 338
  level, 338
tri-state buffer, 299
true rms, 17
truth table, 308
TTL, 314
turn-off delay time, 236
turn-on delay time, 236
turns ratio, 100
twisting, 366
two-integrator oscillator, 260
two-port, 53, 64

U

uncertainty, 20, 358
uncorrelated, 74
unity feedback, 186
unity gain bandwidth, 195
up-counter, 325

V

valance band, 130
value
  average, 29
  mean, 29
varactor, 285
variable
  acrosss-, 43
  complex, 40, 50
  I-, 43
  imaginary, 50
  Laplace, 54
  real, 50
  stochastic, 27
  through-, 43
  time, 52
  V-, 43
variance, 30, 32, 359, 361, 362
VCF, 345
VCO, 266, 282, 343, 346
velocity, 105
Venn diagram, 312
virtual ground, 184
virtual instrument, 348, 354
virtual instrumentation, 354
voltage comparator, 215
voltage controlled filter, 345
voltage controlled oscillator, 266, 282, 343
voltage controlled resistance, 241
voltage divider, 237
voltage generator, 261
voltage limiter, 219
voltage matching, 68
voltage stabilization, 134
voltage stabilizer, 143
voltage switch, 235
voltage-to-current converter, 152
voltage-to-frequency converter, 267

W

weber, 97
Wheatstone bridge, 279, 366
white noise, 75, 344
Wien oscillator, 258
wired-OR, 350

Z

Zener breakdown, 134
Zener diode, 134, 137, 143, 295
Zener reference, 343
Zener voltage, 134
zero, 58
zero drift, 22

Textbooks on science and technology from VSSD

Business engineering, innovation, economics

Creative Facilitation
A Delft approach
Marc Tassoul
2005 / viii+144 p. / ISBN 90-71301-46-X
http://www.vssd.nl/hlf/b005.htm

Fundamentals of Business Engineering and Management
A systems approach to people and organisations
W. Ten Haaf, H. Bikker, D.J. Adriaanse
with contributions from J. in 't Veld and P.Ch-A. Malotaux
2002 / xvi + 728 pp. / ISBN 90-407-2210-2
http://www.vssd.nl/hlf/b001.htm

Macroeconomics
James K. Galbraith and William Darity, Jr.
2005 / xxii+505 p. / ISBN 90-71301-57-5
http://www.vssd.nl/hlf/b006.htm

Civil engineering, coastal engineering

Breakwaters and closure dams
Bed, Bank and Shore Protection 2
K. d'Angremond en F.C. van Roode
2001 / xii+340 p. p. / ISBN 90-407-2127-0
http://www.vssd.nl/hlf/f011.htm

Ebb and Flood Channel Systems in the Netherlands Tidal Waters
Johan van Veen
2002 / 32 p. / ISBN 90-407-2338-9
http://www.vssd.nl/hlf/f015.htm

Introduction to bed, bank and shore protection
Engineering the interface of soil and water
Gerrit J. Schiereck
2004 / xii+ 400 p. / ISBN 90-407-1683-8
http://www.vssd.nl/hlf/f007.htm

Principles of river engineering
The non-tidal alluvial river
P.Ph. Jansen (ed.)
1994 / XVI+509 p. / A4 / ISBN 90-407-1280-8
http://www.vssd.nl/hlf/f006.htm

The vertical motion of foundations and pontoons
Godfried Kruijtzer
2002 / 49 p. / ISBN 90-71301-64-8
http://www.vssd.nl/hlf/f016.htm

Chemical engineering

Mass Transfer in Multicomponent Mixtures
J.A. Wesselingh and R. Krishna
2006 / 329 p. / ISBN 90-71301-58-3
http://www.vssd.nl/hlf/d004.htm

Transport phenomena data companion
L.P.B.M. Janssen and
M.M.C.G. Warmoeskerken
2006 / 168 p. / ISBN 90-71301-59-1
http://www.vssd.nl/hlf/c017.htm

Electronics, physics

Design of High-performance Negative-feedback Amplifiers
E.H. Nordholt
2001 / xvi+234 pag., ISBN 90-407-1247-6
http://www.vssd.nl/hlf/e022.htm

Electronic instrumentation
P.P.L. Regtien
2005 / xiv + 397 p. / ISBN 90-71301-43-5
http://www.vssd.nl/hlf/e008.htm

Electromagnetic Waves — An introductory course
P.M. van den Berg and H. Blok
2001 / X+244 p. / ISBN 90-407-1836-9
includes (though not separately available)
Electromagnetic Waves — A repetitive guide
P.M. van den Berg and M.D. Verweij
1999 / VI+30 p. / ISBN 90-407-1837-7
http://www.vssd.nl/hlf/e016.htm

Materials science

Fracture Mechanics
M. Janssen, J. Zuidema, R.J.H. Wanhill
2002 / xii + 365 p. / ISBN 90-407-2221-8
http://www.vssd.nl/hlf/m004.htm

From Polymers to Plastics
A.K. van der Vegt
2006 / 268 p. / ISBN 90-71301-62-1
http://www.vssd.nl/hlf/m028.htm

Mathematical Geodesy

Series on Mathematical Geodesy and Positioning
http://www.vssd.nl/hlf/a030.htm
Adjustment Theory
P.J.G. Teunissen
2003 / vi + 201 p. / ISBN 90-407-1974-8

Hydrography
C.D. de Jong, G. Lachapelle, S. Skone, I.A.Elema,
2003 / x+351 pp. / ISBN 90-407-2359-1

Dynamic Data Processing
P.J.G. Teunissen
2001 / x + 241 p. / ISBN 90-407-1976-4

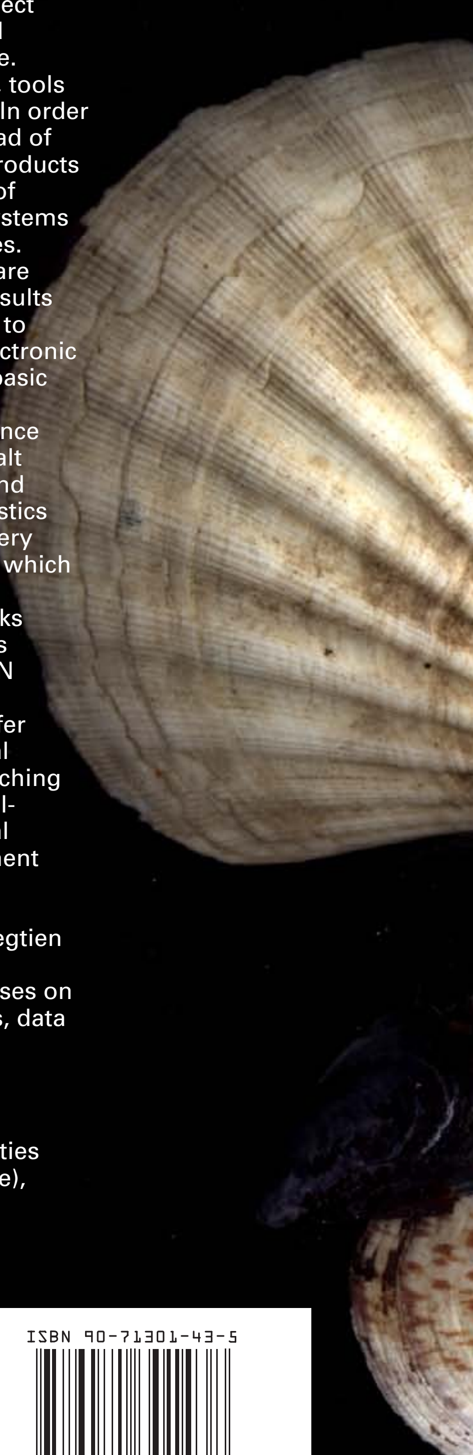
Testing Theory
P.J.G. Teunissen
2000 / viii + 147 p. / ISBN 90-407-1975-6

Mathematics

Mathematical Systems Theory
G.J. Olsder and J.W. van der Woude
2008 / x+208 p. / ISBN 90-71301-40-0
http://www.vssd.nl/hlf/a003.htm

Numerical Methods in Scientific Computing
J. van Kan, A. Segal and F. Vermolen
2005 / xii+277 pp. / ISBN 90-71301-50-8
http://www.vssd.nl/hlf/a002.htm

Order in Space
A.K. van der Vegt
2006 / 93 pp. / ISBN 90-71301-61-3
http://www.vssd.nl/hlf/a017.htm



Electronic instrumentation

Electronic systems have made deep inroads into every aspect of daily life. One need only look around homes, offices and industrial plants to see that they feature almost everywhere. Indeed, it is practically impossible to name any appliances, tools or instruments that do not contain electronic components. In order to compete with rival companies or just remain a step ahead of them, the designers of technical systems and innovative products must be fully aware of both the assets and the limitations of electronic components and systems. Users of electronic systems also need to have a basic knowledge of electronic principles.
In order to fully exploit an instrument's potential, to be aware of its limitations, to correctly interpret the measurement results and to be able to arrive at well-balanced decisions relating to the purchasing, repairing, expansion or replacement of electronic equipment, all users of such systems also need to have a basic knowledge of electronic principles.
This book offers such basic knowledge and provides guidance on how to obtain the relevant skills. The kinds of topics dealt with are operating principles, the performance of analog and digital components and circuits, and the precise characteristics of electronic measuring systems. Throughout the book, every endeavor is made to impart a critical attitude to the way in which such instruments should be implemented.
Contents: 1. Measurement systems • 2. Signals • 3. Networks • 4. Mathematical tools • 5. Models • 6. Frequency diagrams • 7. Passive electronic components • 8. Passive filters • 9. PN diodes • 10. Bipolar transistors • 11. Fieldeffect transistors • 12. Operational amplifiers • 13. Frequency-selective transfer functions with operational amplifiers • 14. Non-linear signal processing with operational amplifiers • 15. Electronic switching circuits • 16. Signal generation • 17. Modulation • 18. Digital-to-analogue and analogue-to-digital conversion • 19. Digital electronics • 20. Measurement instruments • 21. Measurement errors • Appendix • Answers to exercises • Index

About the author: at Delft University of Technology Paul Regtien managed several research projects on robotic sensors and instrumentation. Educational activities were courses on measurement science, measurement transducers, data acquisition and mechatronics.
In 1994 he became a full professor at the Twente University, Faculty of Electrical Engineering, The Netherlands, and head of the Laboratory for Measurement and Instrumentation. Present research activities are measurement science, imaging (optical, acoustic, tactile), mechatronics, sensor technology and instrumentation.

URL about this book: **http://www.vssd.nl/hlf/e008.htm**

Published by VSSD

ISBN 90-71301-43-5 EAN 9789071301438

ISBN 90-71301-43-5

9 789071 301438

e008